(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 July 2003 (24.07.2003)

PCT

(10) International Publication Number
WO 03/059926 A2

(51) International Patent Classification[7]: C07H 21/00

(21) International Application Number: PCT/US02/41064

(22) International Filing Date:
20 December 2002 (20.12.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/343,353 21 December 2001 (21.12.2001) US

(71) Applicant *(for all designated States except US)*: **UNIVERSITY OF MASSACHUSETTS** [US/US]; 365 Plantation Street, Worcester, MA 01605 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **ZHANG, Biliang** [CN/US]; 19 Faulkner Road, Shrewsbury, MA 01545 (US). **CUI, Zhiyong** [CN/US]; 65 Frank Street, Apartment #145, Worcester, MA 01604 (US).

(74) Agents: **GEARY, William, C., III** et al.; Nutter McClennen & Fish LLP, World Trade Center West, 155 Seaport Boulevard, Boston, MA 02210-2604 (US).

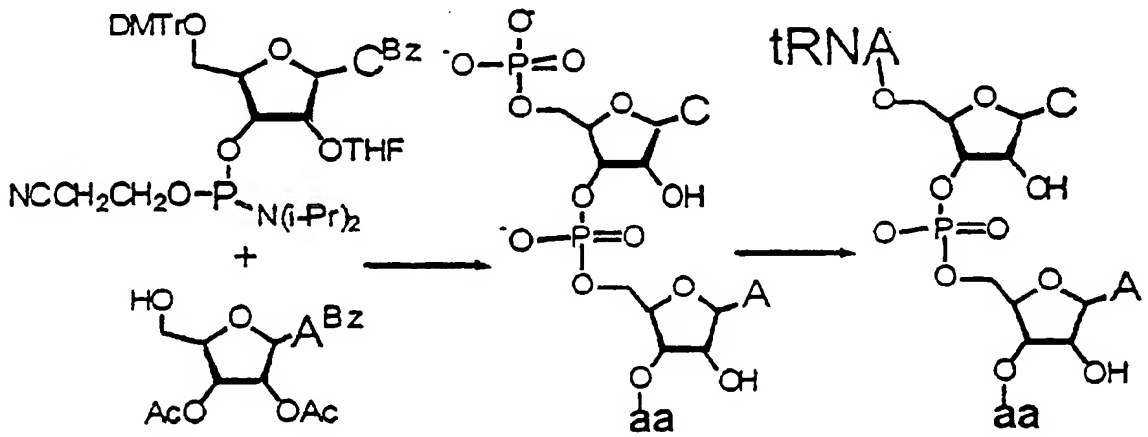
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: METHODS AND COMPOSITIONS FOR AMINOACYL-tRNA SYNTHESIS



(57) Abstract: Methods and compositions are disclosed for the high yield chemical synthesis of 2'(3')-*O*-aminoacylated oligonucleotides, 2'(3')-*O*-aminoacylated pCpA derivatives and 2'(3')-aminoacyl-tRNA's. The present invention discloses the use of tetrahydrofuranyl group (THF) as a stable protecting group in the production of aminoacyl-oligonucleotides. THF can also be used in conjunction with a removable protecting group, such as dimethoxytrityl group (DMTr). The mild conditions employed for the removal of the THF group are compatible with the integrity of the aminoacyl linkage as well as tRNA, which makes it possible to utilize the methods of the present invention to form aminoacyl-tRNA's. The present invention discloses an efficient route for synthesizing 2'(3')-aminoacyl-pCpA, which can be used to load any natural amino acid, unnatural amino acid, labeled amino acid, reporter group or derivative thereof onto a tRNA molecule through an aminoacyl linkage.

WO 03/059926 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHODS AND COMPOSITIONS FOR AMINOACYL-tRNA SYNTHESIS

Priority

The present invention claims priority to U.S. Provisional Application No. 60/343,353, filed December 21, 2001 entitled "Methods and Compositions for Aminoacyl-tRNA Synthesis."

Background of the Invention

The ability to efficiently incorporate natural and unnatural amino acids into proteins will provide new tools to study protein structure and cellular function, and will generate proteins with enhanced properties. Despite decades of intensive research, the catalytic mechanism of protein synthesis in the ribosome is still largely unknown. The peptidyl transferase reaction in the ribosome has never been well characterized because studies have been hampered by technical problems associated with the complexity of the ribosome and its substrates. The 2'(3')-*O*-aminoacyl-pCpA derivatives are the universally conserved terminal sequences of aminoacyl-tRNA. The enzymatic preparation of aminoacyl-tRNA yields only very limited quantities and is not used as a general approach to all natural (see *e.g.*, Biochemistry, 2nd ed, Ed. by L. Stryer, WH Freeman and Co.:1981) and unnatural amino acids (e.g., modified amino acids, such as detectably labeled amino acids, see e.g., Noren et al. (1989) *Science* 244:182-188.) Furthermore, many biophysical studies, such as X-ray crystallography, Fourier transform infrared (FTIR) difference spectroscopy, circular dichroism, and NMR require large quantities of aminoacyl-tRNA's. Thus, the development of general chemical synthesis of 2'(3')-*O*-aminoacyl oligonucleotides is considerably important.

In recent years, the elaboration of proteins containing amino acids and synthetic amino acids at predetermined sites has become technically feasible; however, existing chemical synthesis protocols have numerous drawbacks. Hecht and his colleagues (Hecht, S. M. (1991) *Acc. Chem. Res.* 25, 545) have developed a method involving the coupling of tRNA, missing the 3'-terminal pCpA, with 2'(3')-*O*-aminoacylated pCpA derivatives in the presence of T4 RNA ligase. However, their synthesis of aminoacylated pCpA suffered several drawbacks leading to low yields. Several groups have reported the synthesis of 2'(3')-*O*-aminoacyl-oligonucleotides through either non-

chemical methods or various protective groups and aminoacylating methods. These methods have low aminoacylation yields, long reaction steps, and/or racemization of the products. Schultz and co-workers (Robertson et al.(1991) *J. Am. Chem.* 113: 2722) reported that the aminoacylation of amino acid cyanomethyl esters with unprotected pdCpA was an efficient method for preparing aminoacyl-pdCpA. Although the 2'-deoxycytidine may not be important for the *in vitro* translation system to incorporate a non-natural amino acid into a protein from a suppressor tRNA, it may be considerably critical for studying the peptidyl transferase reaction in the ribosome. Therefore, utilizing aminoacyl-pdCpA may produce questionable *in vivo* information. Literature searches revealed the 2'-OH of nucleotide was either not protected during aminoacylation or was protected with tetrahydropyranyl (THP), 4-methoxy tetrahydropyranyl (Mthp), and fluoride-labile group. These methods were not satisfactory as they resulted in diacylation of the nucleotide, a missing 5'-terminal phosphate that is required for the ligation with tRNA(-CA), extra deprotection steps, or removal of the 2'-OH protecting groups in relatively strong acidic conditions. Existing chemical methods for the chemical aminoacylation of tRNA have many shortcomings, which currently make chemical synthesis of aminoacyl-tRNA inefficient.

Accordingly, there exists a need for new methods and compositions that provide improved chemical synthesis of 2'(3')-*O*-aminoacylated oligonucleotides, 2'(3')-*O*-aminoacyl-pCpA derivatives, and 2'(3')-aminoacyl-tRNA's.

Summary of the Invention

Methods and compositions are disclosed for the high yield chemical synthesis of 2'(3')-*O*-aminoacylated oligonucleotides, 2'(3')-*O*-aminoacylated pCpA derivatives and 2'(3')-aminoacyl-tRNA's. The present invention discloses the use of tetrahydrofuranyl group (THF) as a stable protecting group in the production of aminoacyl-oligonucleotides. THF can also be used in conjunction with a removable protecting group, such as dimethoxytrityl group (DMTr). The mild conditions employed for the removal of the THF group are compatible with the integrity of the aminoacyl linkage as well as tRNA, which makes it possible to utilize the methods of the present invention to form aminoacyl-tRNA's. The present invention discloses an efficient route for synthesizing 2'(3')-aminoacyl-pCpA, which can be used to load any natural amino acid,

unnatural amino acid, labeled amino acid, reporter group or derivative thereof onto a tRNA molecule through an aminoacyl linkage.

The present invention is based, in part, on the discovery that the tetrahydrofuranyl protective group (THF) is compatible with the DMTr group and can be removed during purification allowing for high aminoacylation yield. In one embodiment, DMTr can be selectively removed by catalytic hydrogenation in methanol under atmosphere hydrogen without affecting the THF protecting group. The tetrahydrofuranyl group can serve as a permanent protection group for the 2'-OH of cytidine while 5'-OH can be temporarily protected by the DMTr group. The THF protecting group can later be removed under mild conditions preserving the aminoacyl linkage. In one embodiment, THF is removed by decreasing the pH to between 2 and 5. In another embodiment, THF is removed by acidifying the pH to between 3 and 4.5. In another embodiment, the pH is acidified using a mild acid, *e.g.*, acetic acid.

The present invention pertains to methods for forming aminoacyl-oligonucleotides and aminoacyl-tRNA's. The method comprises the steps of selecting a first nucleotide, protecting its 2'-hydroxyl group with a tetrahydrofuranyl (THF) group and its 5-hydroxyl group with a protecting group that can be selectively removed, selecting a second nucleotide and coupling the first and second nucleotides to form a protected dinucleotide. In one embodiment, the fully protected dinucleotide can be synthesized by phosphoramidite chemistry and treated with ammonium hydroxide to remove the 2-cyanoethyl and benzoyl groups. Following coupling, the 5-hydroxyl group of the dinucleotide can be selectively deprotected using conditions that will not remove THF. The dinucleotide can then be reacted with an aminoacyl-cyanomethyl ester to form a THF protected 2'(3')-O-aminoacyl-oligonucleotide. Any molecule, such as amino acids, unnatural amino acids, reporter groups, or derivatives thereof, may be converted to aminoacyl-cyanomethyl esters and reacted with the dinucleotide. The THF group is then removed, under conditions that will preserve the integrity of the aminoacyl bond, forming a 2'(3')-O-aminoacyl-oligonucleotide. In one embodiment, the nucleotide is selected from the group consisting of guanidine, adenosine, thymidine, cytidine, and uridine. In a preferred embodiment, the first nucleotide is cytidine and the second nucleotide is adenosine.

In another embodiment, the method comprises the steps of chemically synthesizing the dinucleotide pCpA and chemically binding an amino acid, unnatural amino acid or amino acid derivative to the 3'-terminus of pCpA by an ester bond to obtain aminoacyl-pCpA. In general the synthetic method of preparing pCpA derivatives comprises protecting the hydroxyl groups of the nucleotides, independently synthesizing a protected cytidine synthon (7) and adensosine synthon (11), and coupling the two synthons to form a fully protected dinucleotide pCpA (14). In one embodiment, phosphoramidite chemistry is used to couple the two synthons. In one embodiment, the 2'OH of cytidine is protected with a tetrahydrofuranyl group (THF) wherein THF serves as a permanent protection group for the 2'-OH of cytidine. In another embodiment, the 5'-OH of cytidine is protected with a protecting group selected from the group consisting of dimethoxytrityl (DMTr), levulinoyl, and 4, 4' 4"- tris (4,5-dichlorophthalimido) trityl groups, tert-butyldiphenylsilyl (TBDPS), and silyl ether groups. In a preferred embodiment, a DMTr group protects the 5'-OH of cytidine.

The present invention also discloses improved chemical synthesis of aminoacyl-tRNA's. The fully protected dinucleotide pCpA (14) can be converted to a tetrabuylammonium salt (15) and coupled to amino acids and/or non-amino acids through an amino acyl linkage. Amino acids and unnatural amino acids can be chemically coupled to the 3'terminus of the pCpA through an ester bond to form aminoacyl-pCpA's. In one embodiment, dicyclohexylcarbodiimide (DCC) coupling can be used to protect the N-terminus of the amino acids and unnatural amino acids prior to coupling with pCpA. In a preferred embodiment, N-(4-pentenoyl)-amino acids were formed. Following conversion to cyanomethyl esters, N-blocked amino acids and non-amino acids can be coupled to the pCpA dinucleotide. In one embodiment, the N-(4-pentenoyl)-aminoacyl-pCpA can be deprotected and purified. In one embodiment, 2'-O-THF-pCpA can be purified with acetic acid buffer resulting in high yield production of the purified, deprotected 2'(3')-aminoacyl-pCpA product. In another embodiment, N-blocked aminoacyl-pCpA's (21) are produced in high yields.

In one embodiment, aminoacyl-tRNAs can be formed from the chemically synthesized aminoacyl-pCpA. The methods and compositions of the present invention can be used to selective charge tRNA using an RNA ligase. Since T4 RNA ligase is a readily available and inexpensive enzyme, a relatively large quantity of aminoacyl-

tRNA can be prepared by this strategy. A tRNA molecule can be truncated using techniques known in the art to obtain tRNA-C-OH (tRNA(-CA)) lacking pCpA at its 3'-terminus. Subsequently, the aminoacyl-pCpA is enzymatically linked to the 3'-terminus of tRNA-C-3'-OH using RNA ligase to obtain aminoacyl-tRNA. In one embodiment, the dinucleotide protecting groups can be removed following ligation. The improved methods of chemical synthesis described in the present invention allow for deprotection with mild conditions that will preserve the fidelity of the aminoacyl-tRNA. In one embodiment, the cognate amino acid is loaded onto a tRNA molecule. In one embodiment, the tRNA molecule is an initiator tRNA molecule. In another embodiment, the tRNA molecule is a suppresor tRNA molecule.

In another embodiment, the methods of the present invention can be used to misaminoacylate tRNA molecules. A different bond is involved in misaminoacylation than the bond involved with activation of tRNA by aminoacyl tRNA synthetase. Since T4 RNA ligase does not recognize the acyl substituent, tRNA molecules can be readily misaminoacylated with few chemical complications or side reactions. This process is insensitive to the nature of the attached amino acid and allows for misaminoacylation using a variety of unnatural amino acids. In contrast, purely enzymatic aminoacylation is highly sensitive and specific for the structures of substrate tRNA and amino acids. Thus, the improved chemical synthesis of aminoacyl tRNA described in the present invention allows for attachment of any acyl substrate. In one embodiment, the present invention is used to misaminoacylate tRNA with non-cognate amino acids, reporter groups, unnatural amino acids, or derivatives thereof. Natural amino acids, unnatural amino acids, stable isotope labels, and reporter groups (e.g., fluorescent moieties, biotin tags) can be chemically attached to the pCpA dinucleotide and then ligated to a tRNA(-CA) molecule.

In yet another embodiment, the methods of the present invention can be used to form a protein or peptide. Synthesis of 2'(3')-*O*-aminoacylated pCpA derivatives can be used to efficiently incorporate natural amino acids, unnatural amino acids, stable isotope labels, and reporter groups (e.g., fluorescent moieties, biotin tags) into a protein or peptide. The protein can be selected from, but not limited to, recombinant gene products, gene fusion products, enzymes, cytokines, carbohydrate and lipid binding proteins, nucleic acid binding proteins, hormones, immunogenic proteins, human

proteins, viral proteins, bacterial proteins, parasitic proteins and fragments and combinations thereof. In one embodiment, the nascent protein detected is functionally active. In yet another embodiment, the protein contains native amino acids, non-native amino acids, and reporter groups or combinations thereof. Site-directed incorporation of the unnatural amino acids into the protein during translation is not required. Incorporation of non-homologous amino acids can occur anywhere in the polypeptide and can also occur at multiple locations.

The improved chemical synthesis of aminoacyl-pCpA's and aminoacyl-tRNA's disclosed in the present invention can be used in protein engineering, where artificial genes can be developed that encode protein-based materials with desired features. The possibility of efficiently incorporating unnatural amino acids, as well as reporter groups, increases the versatility of protein engineering. In addition, the methods of the present invention can be used to more efficiently study translation and the specificity of the ribosome and may lead to drug discovery.

Brief Description of the Drawings

Figure 1 is a schematic depicting the overall synthesis of 2'(3')-aminoacyl-tRNA from adenosine and THF protected cytidine;

Figure 2 is a schematic depicting the synthesis of 5'-*O*-*tert*-butyldiphenylsilyl-2'-*O*-tetrahydrofuranyl-*N*-benzoyladensine-3'-*O*-[(*N,N*-diisopropyl)-(2-cyanoehtyl)-phosphoramidite];

Figure 3 is a schematic depicting the synthesis of 5'-phosphoryl-2'-*O*-tetrafuranylcytidylyl-(3'→5'phosphoryl) adenosine tris(tetrabutylammonium) salt;

Figure 4 is a schematic depicting the synthesis of 2' (3')-*O*-aminoacylated pCpA and tRNA;

Figure 5 is an autoradiogram of the streptavidin gel-shift analysis of ligation products [2'(3')-aminoacyl-pCpA and 5'-^{32}P-tRNA(-CA)^{Phe}].

BNSDOCID: <WO_____03059926A2_I_>

Figure 6 depicts the ribosomal reaction of the peptide bond formation with chemically synthesized tRNA-Met-Biotin (25) and ^{32}P-CCA-NH-Phe; and

Figure 7 is an autoradiograph demonstrating the peptidyl transferase activity of tRNA-Met-Biotin (25) in the 70S ribosome.

Detailed Description of the Invention

The present invention provides methods and compositions for chemically synthesizing pCpA-2'(3')-aminoacyl derviatives. The invention provides an efficient method for charging tRNA with any compound, including but not limited to, amino acids, unnatural amino acids, labeled amino acids, reporter groups and derivatives thereof. Use of the invention to misaminoacylate tRNA can be employed to study RNA and protein synthesis. The practice of the present invention employs, unless otherwise indicated, conventional methods of organic chemistry, biochemistry, and molecular biology.

So that the invention is more clearly understood, the following terms are defined:

The term "unnatural amino acid" as used herein refers to any compound that can be attached through an acyl linkage. Unnatural amino acid is intended to mean any derivative, analog, or synthetic mimetic of the 20 amino acids or any compound that can be chemically linked through an acyl linkage. The term is intended to cover cyclic derivatives, isomers, and diasteromers as well as the addition of labels to amino acids or amino acid derivatives. Non-limiting examples of unnatural amino acids, include racemic mixtures of selenocysteine and selenomethionine, the D or L forms of nor-leucine, para-nitrophenylalanine, homophenylalanine, para-fluorophenylalanine, 3-amino-p2-benzylpropionic acid, homoarginine, and D-phenylalanine, as well as homologs of natural amino acids, such as ornithine (Orn), norleucine (Nle), pyridylalanine (PAL), γ-(2-naphthyl)-D-alanine (γ -D-2NAL), Nε -5'-(3'-amino-1H-1',2',4'-triazolyl) lysine (Lys(atz)), and the like.

The terms "reporter group" and "marker," as used interchangeably herein, refer to a molecule which can be detected. Preferable the reporter group will be recognized by the enzymes of the translation process and be transferred from a charged tRNA into a growing peptide chain. Reporter groups must also possess certain physical and physio-

chemical properties. Useful physical properties include a characteristic electromagnetic spectral property such as emission or absorbance, ferromagnetism, diamagnetism, magnetism, paramagnetism, luminescence, electrochemiluminescence, fluorescence, phosphorescence, electron spin resonance, electrical capacitance, dielectric constant or electrical conductivity. Many modified amino acids, which can be used as reporter groups, are commercially available (Sigma Chemical; St. Louis, MO; Molecular Probes; Eugene, OR). Alternatively, native amino acids coupled with a detectable label, detectable non-native amino acids, detectable amino acid derivatives and/or analogs can be used.

The reporter group can be used to provide information about the protein structure and conformational changes. The present invention is not intended to be limited by the reporter group. Examples of useful fluorescent moieties that can be used as a reporter group include, but are not limited to, dansyl fluorophores, coumarins and coumarin derivatives, fluorescent acridinium moieties and benzopyrene based fluorophores. Reporter groups include, but are not limited to, biotin, photocleavable biotin, affinity markers (e.g., His 6 tag, C-myc), fluorescent amino acid derivatives (*e.g.*, 5-hydroxyltryptophan, 7-azatryptophan, and ε-dansyllysine), spin –labeled, isotopically labeled, and photoactive (e.g., nitrobenzylester groups) amino acid derivatives or analogs. For example, irradiation of 2-(nitrophenyl)glycine, a non-native amino acid which can be incorporated into the protein using the methods of the present invention, causes site-specific nitrobenzyl-induced photochemical proteolysis. A variety of fluorescent compounds can be used, including fluorescent compounds that have been derivatized (e.g. with NHS) to be soluble (e.g. NHS-derivatives of coumarin). Several BODIPY compounds and reagents have been empirically found to have the additional important and unusual property that they can be incorporated with high efficiency into nascent proteins for both UV and visible excited fluorescence detection. Reporter groups can be used to detect, isolate and quantitate such nascent proteins as recombinant gene products, gene fusion products, truncated proteins caused by mutations in human genes, enzymes, cytokines, hormones, immunogenic proteins, human proteins, carbohydrate and lipid binding proteins, nucleic acid binding proteins, viral proteins, bacterial proteins, parasitic proteins and fragments and combinations thereof.

The term "translation system" as used herein is intended to refer to any system, *e.g.*, cellular or cell-free systems, that supports the incorporation of amino acids into a protein or peptide under the direction of a sequence of nucleotides. A cell-free protein synthesis system consists of mRNA or cDNA encoding for the protein of interest and a reaction mixture comprising the components necessary for translation (e.g., eukaryotic lysates) or transcription/translation (prokaryotic lysates), such as RNA polymerases, tRNAs, aminoacyl synthestases, ribosomes and initiations factors as well as nonrecycled components, *e.g.*, amino acids and nucleotides triphosphates (Zubay, G. (1973) *Annu Rev Genet* 7: 267-287). The cell-free translation system can be selected from the group consisting of Escherichia coli lysates, wheat germ extracts, insect cell lysates, rabbit reticulocyte lysates, frog oocyte lysates, dog pancreatic lysates, human cell lysates, mixtures of purified or semi-purified translation factors and combinations thereof. A cellular translation system includes, but is not limited to, tissue culture cells, primary cells, cells *in vivo*, isolated immortalized cells, human cells and combinations thereof. Cellular translation systems include whole cell preparations such as permeabilized cells or cell cultures wherein a desired nucleic acid sequence can be transcribed to mRNA and the mRNA translated.

I. Translation

Producing proteins from amino acids is a critical process of all living systems. Proteins are formed through the linkage of multiple amino acids via peptide bonds. Key molecules involved in this process are messenger RNA (mRNA) molecules, transfer RNA (tRNA) molecules and ribosomes (rRNA-protein complexes). Protein translation normally occurs in living cells and in some cases can also be performed outside the cell in systems referred to as cell-free translation systems. In either system, the basic process of protein synthesis is identical. The extra-cellular or cell-free translation system comprises an extract prepared from the intracellular contents of cells. These preparations contain those molecules which support protein translation and depending on the method of preparation, post-translational events such as glycosylation and cleavages as well. Typical cells from which cell-free extracts or *in vitro* extracts are made are *Escherichia coli* cells, wheat germ cells, rabbit reticulocytes, insect cells and frog oocytes. In a preferred embodiment, *Escherichia coli* cell extracts are used.

During translation, specific tRNAs (transfer RNA) pick up specific amino acids, transfer those amino acids to the ribosomes, and insert them in their proper place according to the mRNA is done by the anticodon portion of the tRNA molecules complementary base pairing with the codons along the mRNA. Transfer RNA (tRNA) is a three-dimensional, inverted cloverleaf-shaped molecule of RNA about 70 nucleotides long. At the top, or 3' end, a specific amino acid can be attached to a specific tRNA, wither through specific enzymes called aminoacyl-tRNA synthetases or chemical synthesis as discussed in this invention. The resulting complex of an amino acid and a tRNA is referred to as an aminoacyl-tRNA.

At the bottom loop of the cloverleaf is a series of three unpaired tRNA bases called the anticodon. An anticodon is a series of three tRNA bases complementary to a mRNA codon. One loop of the tRNA binds to the aminoacyl-tRNA synthetase charging enzyme while the other loop attaches to the 50S ribosomal subunit. Most of the remaining tRNA bases are involved in intrastrand hydrogen bonds that gives the tRNA its specific shape. While there are 64 different mRNA codons, there are not 64 different molecules of tRNA. For example, there are no tRNA molecules that possess an anticodon complementary to the three nonsense or stop codons. Furthermore, the anticodons of some tRNAs are able to recognize more than one codon because the tRNA's recognition of the third nucleotide of the codon is not always precise. However, the right amino acid is still inserted because there are 61 codons that code for the 20 different amino acids.

A number of protein factors associate with the ribosome during different phases of translation including initiation factors, elongation factors and termination factors. The ribosomal subunits in eukaryotic cells have a density of 60S and 40S while those in prokaryotes have a density of 50S and 30S. Translation begins with the binding of the ribosome to mRNA. To initiate translation in prokaryotic cells, a 30S ribosomal subunit binds to a short nucleotide sequence on the mRNA called the ribosome binding site. However, translation does not usually begin until the 30S ribosomal subunit reaches the first AUG sequence in the mRNA. For this reason, AUG is known as the start codon. At this point, an initiation complex composed of the 30S subunit, a tRNA having the anticodon UAC and carrying an altered form of the amino acid methionine (*N*-formylmethionine or f-Met), and proteins called initiation factors is formed. A 50S

ribosomal subunit then attaches to the initiation complex and the initiation factors leave. This forms the 70S ribosome. The joining of individual amino acids to form a protein or polypeptide is known as the elongation phase of translation. There are two sites on the 70S ribosome. The A or acceptor or aminoacyl site is where an aminoacyl-tRNA first attaches. The P or peptide site is where the growing amino acid chain is temporarily being held by a tRNA as the next codon in the mRNA is being read. During peptide bond formation, the amino acid chain or peptide moves from the tRNA at the P-site and forms a peptide bond with the new amino acid attached to the tRNA at the A-site. The peptide bond is formed by a ribozyme, an enzyme composed of RNA, called peptidyl transferase. The now uncharged tRNA at the P-site leaves the ribosome through an adjacent portion called the E-site (exit site) to eventually pick up a new amino acid and be recycled. Meanwhile, the 70S ribosome moves a distance of one codon down the mRNA through a process called translocation to allow decoding of the next codon in the message. The growing polypeptide chain actually passes through a tunnel in the 50S ribosomal subunit. This process continues over and over again in the 5' to 3' direction until the ribosome hits a stop codon. A stop codon is a series of three mRNA bases coding for no amino acid and thus terminates the protein chain. UAA, UAG, UGA are the three stop codons in the genetic code. Stop codons do not code for an amino acid because they cannot be recognized by a tRNA. Proteins called release factors free the protein from the tRNA and the two ribosomal subunits come apart to be recycled. During this elongation process, the protein has assumed its three-dimensional functional shape (See Alberts et al. Molecular Biology of The Cell, Garland Publishing, Inc. NY).

Aminoacylation or charging of tRNA results in linking the carboxyl terminal of an amino acid to the 2'-(or 3'-) hydroxyl group of a terminal adenosine base via an ester linkage. This process can be accomplished either using enzymatic or chemical methods. Normally a particular tRNA is charged by only one specific native amino acid. This selective charging, enzymatic aminoacylation, is accomplished by aminoacyl tRNA synthetases. A tRNA which selectively incorporates a tyrosine residue into the nascent polypeptide chain by recognizing the tyrosine UAC codon will be charged by tyrosine with a tyrosine-aminoacyl tRNA synthetase, while a tRNA designed to read the UGU codon will be charged by a cysteine-aminoacyl tRNA synthetase. These synthetases have evolved to be extremely accurate in charging a tRNA with the correct amino acid

to maintain the fidelity of the translation process. Except in special cases where the non-native amino acid is very similar structurally to the native amino acid, it is necessary to use means other than enzymatic aminoacylation to charge a tRNA. The present invention describes an improved chemical synthesis for aminoacyl tRNA. A tRNA charged with anything but its native amino acid molecule is referred to as a misaminoacylated tRNA molecule. In one embodiment, the present invention can be use to misaminoacylate tRNA molecules, load a different compound (e.g., unnatural amino acid, amino acid derivative, non-homologous amino acid, reporter group) onto an tRNA molecule. In another embodiment, the methods of this invention can be used to load unnatural amino acids onto tRNA. In yet another embodiment, the methods of this invention can be used to load tRNA with a reporter group.

Typically, radioactively labeled amino acid residues, such as ^{35}S-methionine, can be used as a means of detecting newly synthesized proteins or nascent proteins. These nascent proteins can normally be distinguished from the many other proteins present in a cell or a cell-free extract by first separating the proteins by the standard technique of gel electrophoresis and determining if the proteins contained in the gel possess the specific radioactively labeled amino acids. Prior knowledge of the expressed protein is not required and in general does not require the protein to have any special properties. In addition, the protein can exist in a denatured or unfolded form for detection by gel electrophoresis. Similarly, the charged tRNA can be shown to be a substrate for the ribosome through the incorporation of ^{32}P following a ribosomal reaction between chemically synthesized tRNA-Met-Biotin (25) and ^{32}P-CCA-NH-Phe as depicted in Figure 6 and described in Example 3. Radioactive assays also have the advantage that the structure of the nascent protein is not altered or can be restored, and thus, proteins can be isolated in a functional form for subsequent biochemical and biophysical studies.

Another embodiment of the invention is directed to methods for labeling nascent proteins. An initiator tRNA molecule, such as methionine-initiator tRNA or formylmethionine-initiator tRNA can be misaminoacylated with a fluorescent moeity (e.g. a BODIPY moiety) and introduced to a translation system. The system is incubated and the reporter group can be incorporated at the amino terminus of the nascent proteins. Nascent proteins containing reporter groups can be detected, isolated and quantitated. Reporter groups or parts of reporter groups may be cleaved from the nascent proteins

which substantially retain their native configuration and are functionally active. The present invention can be used with any tRNA molecule, including, but not limited to a tRNA molecule, an initiator tRNA molecule, a suppressor tRNA molecule, as well as genetically engineered tRNAs (*See* (Kowal et al. *Nucleic Acids Res* (1997) 25:4685-4689); (Cload et al. *Chem Biol* (1996) 3: 1033-1038)). Representative sequence files of tRNA molecules from a variety of species, e.g., *E. coli, B. subtilis, S. cerevisiae, A. thaliana, D. melanogaster, H. sapiens,* and *H. volcanii*, can be obtained from the databases of EMBL and GenBank Web sites.

II. Unnatural aminoacids

Incorporating the non-protein amino acids into the growing protein chain can be useful in drug disovery. Of the thousands of known non-proteinogenic amino acids, about 300 occur in plants. They are found mostly in a small number of families, such as the Leguminosae, Cucurbitaceae, Sapindacae, Aceraceae and Hippocastenaceae. Many of these non-amino acids are structurally similar to the components of common proteins. The incorporation of non-amino acids into proteins may be associated with autoimmune diseases in humans. Furthermore, many natural product antiobiotics involve the incorporation of non-proteinogenic amino acids (Chartrain M. et al *Curr Opin Biotechnol* (2000) 11(2):209-214); (Dougherty DA. *Curr Opin Chem Biol* (2000) 4(6):645-652).

Unnatural amino acids can also be used to introduce diversity and structural constraints and enhance biological acitivity and proteolytic resistence of peptide ligand mimetics. Unnatural amino acids are readily available (See *e.g.*, Sigma-Aldrich Co.). In addition, many chemical and enzymatic methods are available to prepare unnatural amino acids ((Ager DJ et al. *Curr Opin Drug Discov Devel* (2001) 4(6):800-807); (Altenbuchner J et al. *Curr Opin Biotechnol* (2001) 12(6):559-563); (Bouifraden S et al. *Amino Acids* (1999)16(3-4):345-379). Unnatural amino acids include, but are not limited to, β-amino acids, homo-amino acids, cyclic amino acids, aromatic amino acids, D-amino acids, alanine derivatives, glycine derivatives, ring-substituted Phe and Tyr derivatives, linear core amino acids, diamino acids, and N-Boc monoprotected diamines.

III. Reporter Groups

Reporter groups or markers are basically molecules which will be recognized by the enzymes of the translation process and transferred from a charged tRNA into a growing peptide chain. To be useful, reporter groups must also possess certain physical and physio-chemical properties. For example, a reporter group must be suitable for incorporation into a growing peptide chain. This may be determined by the presence of chemical groups which will participate in peptide bond formation. The reporter groups should also be attachable to a tRNA molecule. Attachment is a covalent interaction between the 3'-terminus of the tRNA molecule and the carboxy group of the report group or a linking group attached to the reporter group and to a truncated tRNA molecule. Linking groups may be nucleotides, short oligonucleotides or other similar molecules and are preferably dinucleotides and more preferably the dinucleotide CA. The reporter groups should also have one or more physical properties that facilitate detection and possibly isolation of nascent proteins. Useful physical properties include a characteristic electromagnetic spectral property such as emission or absorbance, magnetism, electron spin resonance, electrical capacitance, dielectric constant or electrical conductivity.

Useful reporter groups are native amino acids coupled with a detectable label, detectable non-native amino acids, detectable amino acid analogs and detectable amino acid derivatives. Labels and other detectable moieties may be ferromagnetic, paramagnetic, diamagnetic, luminescent, electrochemiluminescent, fluorescent, phosphorescent, chromatic or have a distinctive mass. Fluorescent moieties which are useful as markers include dansyl fluorophores, coumarins and coumarin derivatives, fluorescent acridinium moieties and benzopyrene based fluorophores. Preferably, the fluorescent marker has a high quantum yield of fluorescence at a wavelength different from native amino acids and more preferably has high quantum yield of fluoresence can be excited in both the UV and visible portion of the spectrum. Upon excitation at a preselected wavelength, the marker is detectable at low concentrations either visually or using conventional fluorescence detection methods. Electrochemiluminescent markers such as ruthenium chelates and its derivatives or nitroxide amino acids and their derivatives are preferred when extreme sensitivity is desired (DiCesare et al.,

BioTechniques (1993) 15:152-159). These markers are detectable at the femtomolar ranges and below.

In addition to fluorescent markers, a variety of markers possessing other specific physical properties can be used to detect nascent protein production. In general, these properties are based on the interaction and response of the marker to electromagnetic fields and radiation and include absorption in the UV, visible and infrared regions of the electromagnetic spectrum, presence of chromophores which are Raman active, and can be further enhanced by resonance Raman spectroscopy, electron spin resonance activity and nuclear magnetic resonances and use of a mass spectrometer to detect presence of a marker with a specific molecular mass. These electromagnetic spectroscopic properties are preferably not possessed by native amino acids or are readily distinguishable from the properties of native amino acids. For example, the amino acid tryptophan absorbs near 290 nm, and has fluorescent emission near 340 nm when excited with light near 290 nm. Thus, tryptophan analogs with absorption and/or fluorescence properties that are sufficiently different from tryptophan can be used to facilitate their detection in proteins.

Reporter groups can be chemically synthesized from a native amino acid and a molecule with marker properties which cannot normally function as an amino acid. For example a highly fluorescent molecule can be chemically linked to a native amino acid group. The chemical modification can occur on the amino acid side-chain, leaving the carboxyl and amino functionalities free to participate in a polypeptide bond formation. For example, highly fluorescent dansyl chloride or coumarin can be linked to the nucleophilic side chains of a variety of amino acids including lysine, arginine, tyrosine, cysteine, histidine, etc., mainly as a sulfonamide for amino groups or sulfate bonds to yield fluorescent derivatives. Such derivatization leaves the ability to form peptide bond intact, allowing the normal incorporation of dansyllysine into a protein.

IV. Chemical Synthesis of 2'(3')-Aminoacylated Oligonucleotides

In one embodiment the present invention pertains to the general synthesis of 2'(3')-aminoacylated oligonucleotides. The methods disclosed in the present invention can be used to form any oligonucleotide. The starting nucleotides can be selected from the group consisting of guanidine, adenosine, thymidine, cytidine, and uridine. The

method comprises the steps of selecting a first nucleotide, protecting its 2'-hydroxyl group with a tetrahydrofuranyl (THF) group and its 5-hydroxyl group with a protecting group that can be selectively removed, selecting a second nucleotide and coupling the first and second nucleotides to form a protected dinucleotide. Following coupling, the 5-hydroxyl group of the dinucleotide can be selectively deprotected using conditions that will not remove THF. The dinucleotide can then be reacted with an aminoacyl-cyanomethyl ester to form a THF protected 2'(3')-O-aminoacyl-oligonucleotide. Any molecule, such as amino acids, unnatural amino acids, reporter groups, or derivatives thereof, may be converted to aminoacyl-cyanomethyl esters and reacted with the dinucleotide. The THF group is then removed, under conditions that will preserve the integrity of the aminoacyl bond, forming a 2'(3')-O-aminoacyl-oligonucleotide. The method can be expanded to synthesize an oligonucleotide of any length via successive coupling reactions. In a preferred embodiment, the first nucleotide is cytidine and the second nucleotide is adenosine. As a non-limiting example of the methods of the present invention, the synthesis of 2'(3')-aminoacylated pCpA is described below. In another embodiment, the synthesized 2'(3')-aminoacylated pCpA can be ligated to a tRNA molecule producing aminoacyl-tRNA as depicted schematically in Figure 1.

A. Synthesis of a Cytidine Synthon

In a preferred embodiment, the 2'(3')-aminoacylated oligonucleotide is 2'(3')-aminoacylated pCpA. This synthesis depends on the formation of a cytidine synthon, 4-*N*-benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-tetrafuranyl -cytidine-3'-[(2-cyanoethyl)-1-(*N,N*-diisopropyl)] phosphoramidite (7), which is blocked at the 3'OH with the protecting group tetrahydrofuranyl (THF). The synthesis of a cytidine synthon is depicted in Figure 2, which illustrates the various reaction steps. Representative reactions forming intermediates and their respective conditions for this embodiment are provided in greater detail below and are intended to be included within the scope of the invention.

In one embodiment, cytidine is used as the starting nucleotide. Cytidine (1) was reacted first with trimethylsilyl chloride in pyridine and then with benzoyl chloride and was finally treated with ammonium hydroxide in the same reaction pot to produce 4-*N*-benzoyl-cytidine (2) in 94% yield. In a preferred embodiment, 4-*N*-benzoyl-cytidine (2)

was produced by adding trimethylchlorosilane (38.1mL, 300 mmol) to a suspension of cytidine (9.72 g, 39.96 mmol) in pyridine (200 mL). After the reaction mixture stirred for 1 h at room temperature under argon atmosphere, benzoyl chloride (23.0 mL, 200 mmol) can be injected and the mixture stirred at room temperature for 4 h. The mixture was then cooled in an ice bath and water (40 mL) was added. After 20 min 29% of aqueous ammonium hydroxide (60 mL, 445.8 mmol) was added and the mixture was stirred at room temperature for 25 min. The reaction was then evaporated to near dryness and the residue is dissolved in water (400 mL). Crystallization began while the solution was washed by EtOAc (150 mL). The solid was filtered out and washed by EtOAc (30 mL), ice water (30 mL), and acetone (30 mL, twice) to give compound (2) (13.1 g, 94.4%) as white solid: R_f = 0.55 (20% methanol in EtOAc); ^{1}H NMR (DMSO-d_6, 400 MHz) δ 3.57-3.77 (m, 2H), 3.89-4.01 (m, 3H), 5.06 (d, J = 5.47 Hz, 1H), 5.18 (t, J = 5.08 Hz, 1H), 5.52 (d, J = 5.09 Hz, 1H), 5.79 (d, J = 2.74 Hz, 1H), 7.31 (m, 1H), 7.48-7.63 (m, 3H), 7.97-7.99 (m, 2H), 8.47 (d, J = 7.42 Hz, 1H), 11.16 (br s, 1H); ^{13}C NMR (DMSO-d_6, 100 MHz) δ 60.57, 69.32, 75.23, 84.90, 90.88, 96.80, 129.14, 133.42, 133.86, 145.94, 155.22, 163.70, 168.03.

The second step in the synthesis of the cytidine synthon involves the reaction of 4-*N*-benzoyl-cytidine (2) with 1,3-dichloro-1,1,3,3-tetraisopropyldisiloxane (TIPDSCl$_2$) in pyridine to yield 4-*N*-benzoyl-3',5'-*O*-(tetraisopropyldisiloxane-1,3-diyl) cytidine (3). In one embodiment, compound (3) was produced in 94% yield. In a preferred embodiment, 4-*N*-benzoyl-3',5'-*O*-(tetraisopropyldisiloxane-1,3-diyl)cytidine (3) was produced by slowly adding1,3-dichloro-1,1,3,3-tetraisopropyl-1,3-disiloxane (5.5 mL, 17.34 mmol) to a suspension of (2) (5.02 g, 14.45 mmol) in anhydrous pyridine (200 mL) at 0 °C. After being stirred 6 hours at room temperature under argon atmosphere, water (4 mL) was added to quench the reaction at 0 °C. After the solvent was removed, the residue was dissolved in CH_2Cl_2 (600 mL) and was washed by water (100 mL) and saturated aqueous NaCl (90 mL). The organic layer was dried over anhydrous Na_2SO_4 and the solvent was removed under reduced pressure. The residue was purified by flash column chromatography (SiO_2, 9.0-60% EtOAc in hexane) to give compound (3) (8 g, 94.0%) as white solid: R_f = 0.57 (9.1% methanol in chloroform); ^{1}H NMR (400 MHz, $CDCl_3$) δ 0.94-1.11 (m, 28H), 2.88 (br s, 1H), 4.01-4.05 (dd, J = 13.44, 2.69 Hz, 1H) 4.20-4.35 (m, 4H), 5.86 (s, 1H), 7.50-7.63 (m, 4H), 7.87-7.89 (d, J = 6.59 Hz, 2H), 8.22

(d, J = 6.60 Hz, 1H), 8.67 (br s, 1H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 12.70, 13.13, 13.15, 13.60, 17.05, 17.12, 17.21, 17.52, 17.63, 17.68, 60.23, 68.82, 75.46, 82.26, 91.80, 96.24, 127.70, 129.34, 133.49, 144.92, 162.56; ESI-MS calculated M 589.2640, found (M + H) 590.2.

The third step in the cytidine synthon is the protection of the 2'OH of cytidine with THF. The present invention discloses the use of tetrahydrofuranyl group as a stable protecting group which can be used in conjunction with selectively removeable 5'OH protecting groups, including, but not limited to, dimethoxytrityl (DMTr), levulinoyl, and 4, 4' 4"- tris (4,5-dichlorophthalimido) trityl groups, tert-butyldiphenylsilyl (TBDPS), and silyl ether groups. In a preferred embodiment, the 5'OH protecting group is the dimethoxytrityl group (DMTr). The 5'OH protecting group can be selectively removed using a variety of methods known in the art. In a preferred embodiment, DMTr can be selectively removed by catalytic hydrogenation in methanol under atmosphere hydrogen without affecting the THF protecting group. The tetrahydrofuranyl group can serve as a permanent protection group for the 2'-OH of cytidine while 5'-hydroxy can be protected by the DMTr group, which can be selectively removed. 2'(3')-aminoacylated oligonucleotides are stable between pH 2 and pH 7, so the ideal deprotection of the 2'-OH protective group of cytidine should be conducted within this range. Tetrahydrofuranyl (THF) group can be removed under these conditions. The compatibility of the THF protective group with the deprotection conditions of regular 5'-OH protective groups like DMTr, levulinoyl, silyl ether groups, and 4, 4', 4"-tris (4,5-dichlorophthalimido) trityl group is disclosed in this invention.

Compound (3) was reacted with 2,3-dihydrofuran in the presence of *p*-toluenesulfonic acid in THF to produce 4-*N*-benzoyl-2'-tetrahydrofuranyl-3',5'-*O*-(tetraisopropyl disiloxane-1,3-diyl) cytidine (4). Comound 3 and *p*-toluenesulfonic acid were dried by oil pump for 5 h. To a solution of compound (3) (12.89 g, 21.87 mmol) and *p*-toluenesulfonic acid (1.46 g, 7.67 mmol) in THF (250 mL), 2,3-dihydrofuran (16.7 mL, 219.2 mmol) is added by syringe at 0 °C. The reaction mixture was stirred for 3 hours at 0°C under argon atmosphere. The concentrated ammonium hydroxide (11.7 mL) was added to reaction mixture, the solvent was removed under reduced pressure at low temperature. The residue was dissolved in $CHCl_3$ (800 mL) and washed by saturated aqueous $NaHCO_3$ (100 mL) and water (100 mL). The aqueous layers was

extracted by $CHCl_3$ (100 mL). The combined organic layers are dried over Na_2SO_4 and the solvent was removed under reduced pressure. The residue was purified by flash column chromatography (SiO_2, 3.3-75% EtOAc in hexane) to give (4) in 97% yield (14 g, 97.0%, two diastereoisomers) as white solid: R_f = 0.67 (75% EtOAc in hexane); ^1H NMR (400 MHz, $CDCl_3$) δ 0.92-1.11 (m, 28H), 1.81-2.00 (m, 4H), 3.82-4.36 (m, 7H), 5.64-5.74 (m, 1H), 5.84 (d, J = 8.99 Hz, 1H), 745-7.61 (m, 4H), 7.88 (d, J = 7.42 Hz, 2H), 8.30 (m, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$) δ 12.76, 12.80, 13.04, 13.23, 13.59, 13.63, 17.06, 17.11, 17.16, 17.18, 17.22, 17.30, 17.46, 17.51, 17.61, 17.68, 22.76, 23.16, 32.23, 32.74, 59.58, 59.64, 66.67, 66.77, 67.21, 68.95, 75.80, 77.80, 81.99, 82.20, 90.71, 90.85, 96.57, 102.61, 103.06, 128.02, 128.99, 133.15, 133.47, 144.34, 154.69, 162.94, 171.34; ESI-MS calculated M 659.3058, found (M + H) 660.2, (M + Na) 682.1.

After introducing the tetrahydrofuranyl (THF) group at the 2'-position, the 3',5'-tetraisopropyldisiloxanyl group was removed by treating with *tetra*-butylammonium fluoride in THF to generate 4-*N*-benzoyl-2'-*O*-tetrahydrofuranylcytidine (5). Tetrabutylammonium fluoride (5.3 mL, 5.24 mmol) was added to a solution of compound (4) (3.46 g, 5.24 mmol) in THF (200 mL) at 0 °C under argon atmosphere. The reaction mixture was stirred for 6 h. A solution of 5% NH_4HCO_3 (6 mL) was added to quench the reaction. After the solvent was removed, the residue was dissolve in a mixed solvent of CH_2Cl_2 and pyridine (600 mL, v/v = 3/1) and washed by saturated aqueous $NaHCO_3$ (50 mL, twice). The organic layer was dried over Na_2SO_4 and the solvent was removed under reduced pressure. The residue was purified by flash column chromatography (SiO_2, 0.1-3.0% CH_3OH in CH_2Cl_2) to give (5a) and (5b) in 90% yield (49% high isomer and 41% low isomer) (1.97 g, 90%; two diastereoisomers, 5a, 1.06 g, 48.7%; 5b, 0.90 g, 41.3%) as white solid. For compound (5a): R_f = 0.46 (9.1% methanol in chloroform); ^1H NMR (400 MHz, $CDCl_3$) δ 1.84-2.05 (m, 4H), 3.58 (d, J = 5.86 Hz, 1H), 3.80-4.09 (m, 6H), 4.30 (m, 1H), 4.46 (m, 1H), 5.47 (m, 1H), 5.78 (d, J = 2.05 Hz, 1H), 7.44-7.56 (m, 4H), 7.86 (m, 2H), 8.40 (d, J = 7.32 Hz, 1H), 9.14 (br s, 1H); ^{13}C NMR (100 MHz, DMSO-d_6) δ 23.51, 32.47, 60.02, 66.92, 68.21, 79.01, 84.95, 89.40, 96.77, 102.87, 129.12, 129.42, 133.42, 133.77, 145.62, 155.05, 163.82, 168.11; ESI-MS calculated M 417.1536, found (M - H) 416.2. For compound (5b): R_f = 0.39 (9.1% methanol in chloroform); ^1H NMR (400 MHz, DMSO-d_6) δ 1.02-1.86 (m, 4H), 3.59-3.67 (m, 2H), 3.88 (m, 1H), 4.11-4.18 (m, 2H), 5.14-5.17 (m, 2H), 5.26 (m, 1H), 5.93 (d,

J = 4.69 Hz, 1H), 7.31-7.61 (m, 4H), 7.99 (d, J = 7.42 Hz, 2H), 8.39 (d, J = 7.04 Hz, 1H), 11.25 (br s, 1H); ^{13}C NMR (100 MHz, DMSO-d_6) δ 23.27, 32.54, 61.19, 66.92, 69.20, 78.49, 85.90, 88.78, 96.97, 103.40, 129.12, 129.16, 133.42, 133.84, 146.28, 155.30, 163.70, 168.15; ESI-MS calculated M 417.1536, found (M - H) 416.3.

The next step in the chemical synthesis of the cytidine synthon is selective protection of the 5'-hydroxyl group. In a preferred embodiment, the 5'-OH is protected with a dimethoxytrityl group (DMTr), *e.g.*, 4,4'-dimethoxytrityl chloride, to yield 4-*N*-benzoyl-2'-*O*-tetrahydrofuranyl-5'-*O*-(4,4'-dimethoxytrityl)cytidine (6). To a solution of (5b) (0.47 g, 1.13 mmol) and 4,4'-dimethoxytrityl chloride (0.763 g, 2.25 mmol) (dried over vacuum pump over night) in pyridine (40 mL), diisopropylethylamine (0.98 mL, 5.63 mmol) is added. The reaction mixture is stirred at room temperature under argon atmosphere for 12 h. EtOH (2 mL) is added and the solvent is removed. The residue was dissolved in $CHCl_3$ (300 mL) and washed by saturated aqueous $NaHCO_3$ (30 mL) and saturated aqueous NaCl (30 mL). The organic layer is dried over Na_2SO_4 and the solvent was removed under reduced pressure. The residue is purified by flash column chromatography (SiO_2, 0.1-0.4% CH_3OH in $CHCl_3$) to give (6b) (0.73 g, 90.1%) as yellow solid: R_f = 0.36 (4.8% methanol in chloroform); ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.85-2.06 (m, 4H), 2.56 (d, J = 7.03 Hz, 1H), 3.48-3.58 (m, 2H), 3.82 (s, 6H), 3.87-3.97 (m, 2H), 4.12-4.14 (m, 1H), 4.36-4.42 (m, 2H), 5.54 (m, 1H), 6.27 (s, 1H), 6.86-6.88 (m, 4H), 7.26-7.63 (m, 13H), 7.86 (m, 2H), 8.35 (br s, 1H), 8.55 (s, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$) δ 23.46, 32.65, 55.47, 62.24, 68.05, 69.14, 79.43, 83.70, 87.29, 89.47, 97.12, 104.59, 113.56, 127.37, 127.98, 128.28, 128.50, 129.19, 130.30, 130.39, 133.27, 135.59, 135.87, 144.41, 145.39, 155.20, 158.89, 162.61; ESI-MS calculated M 719.2843, found (M + H) 720.1, (M + Na) 742.3. Similar treatment of compound (5a) (0.48 g, 1.15 mmol) and 4,4'-dimethoxytrityl chloride (0.779 g, 2.3 mmol) as that described for compound (5b) gave compound (6a) (0.78 g, 94%); R_f = 0.40 (9.1% methanol in chloroform); ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.84-2.08 (m, 4H), 3.55-3.59 (m, 3H), 3.81 (s, 6H), 3.84-4.09 (m, 3H), 4.33-4.43 (m, 2H), 5.56 (m, 1H), 5.94 (s, 1H), 6.86-6.89 (m, 4H), 7.24-7.57 (m, 13H), 7.88-7.91 (m, 2H), 8.54 (d, J = 7.42 Hz, 1H), 8.95 (br s, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$) δ 24.01, 32.78, 55.48, 61.32, 68.04, 68.18, 81.99, 83.40, 87.25, 90.36, 96.57, 105.36, 113.55, 127.39, 127.74, 128.28, 128.51,

129.27, 130.30, 130.39, 133.36, 135.59, 135.90, 144.39, 145.17, 155.08, 158.90, 162.38; ESI-MS calculated, M 719.2843, found (M + H) 720.1, (M + Na) 742.3.

The next step completes the synthesis of the THF protected cytidine synthon. Compound (6) was coupled with chloro(2-cyanoethoxy) (*N,N*-diisopropylamino) phosphine to afford the building block 4-*N*-benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-tetrafuranyl-cytidine-3'-[(2-cyanoethyl)-1-(*N,N*-diisopropyl)] phosphoramidite (7). Compound (7) can be produced in 55-59% yield, R_f = 0.59, 0.70 (4.8% EtOAc in chloroform); ^{1}H NMR ($CDCl_3$) δ1.00-1.15 (m, 12H), 1.76-1.94 (m, 4H), 2.53 (t, J = 6.25 Hz, 1H), 2.61 (t, J = 6.25 Hz, 1H), 3.42-3.66 (m, 6H), 3.77-3.76 (m, 6H), 3.83-3.91 (m, 2H), 4.22-4.30 (m, 1H), 4.44-4.53 (m, 2H), 5.44-5.48 (m, 1H), 6.19-6.21 (m, 1H), 6.81-6.85 (m, 4H), 7.02 (br s, 1H), 7.20-7.54 (m, 13H), 7.88-7.90 (m, 2H), 8.26-8.36 (m, 1H); ^{31}P-NMR ($CDCl_3$) δ150.92, 151.38; ESI-MS calcld. 919.4, found 920.1 $(M + H)^+$. In one embodiment, to an ice bath cooled solution of compound (6b) (0.8 g, 1.11 mmol) dried by oil pump over night previously in CH_2Cl_2 (25 mL), diisopropylethylamine (0.794 mL, 4.56 mmol) and chloro(2-cyanoethoxy)(*N,N*-diisopropylamino)phosphoine are added successively under argon atmosphere. After 20 min the reaction mixture is stirred at room temperature for 4.5 h. The mixture is diluted by EtOAc (250 mL) and the organic layer was washed by saturated aqueous $NaHCO_3$ (25 mL) and saturated aqueous NaCl (25 mL). The organic layer was dried over Na_2SO_4 and the solvent was removed under reduced pressure. The residue was purified by column chromatography (SiO_2, 77% EtOAc/19.2% hexane/3.8% TEA) to give 4-*N*-benzoyl-2'-*O*-tetrahydrofuranyl-5'-*O*-(4,4'-dimethoxytrityl)cytidine-3'-*O*-(2-cyanoethyl *N,N*-diisopropylphosphoramidite (7b) (0.6 g, 58.5%, two diastereomeric spots) as yellow solid: R_f = 0.59, 0.70 (4.8% EtOAc in chloroform); ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.00-1.15 (m, 12H), 1.76-1.94 (m, 4H), 2.53 (t, *J* = 6.25 Hz, 1H), 2.61 (t, *J* = 6.25 Hz, 1H), 3.42-3.66 (m, 6H), 3.77-3.76 (m, 6H), 3.83-3.91 (m, 2H), 4.22-4.30 (m, 1H), 4.44-4.53 (m, 2H), 5.44-5.48 (m, 1H), 6.19-6.21 (m, 1H), 6.81-6.85 (m, 4H), 7.02 (br s, 1H), 7.20-7.54 (m, 13H), 7.88-7.90 (m, 2H), 8.26-8.36 (m, 1H); ^{13}C-NMR (100 MHz, $CDCl_3$) δ 20.36, 20.43, 20.57, 20.63, 23.17, 23.31, 24.70, 24.77, 24.82, 24.86, 32.58, 32.73, 43.27, 43.40, 43.53, 55.41, 55.44, 58.17, 58.37, 58.71, 58.89, 61.83, 62.23, 67.50, 67.54, 70.47, 70.53, 70.61, 70.68, 82.98, 83.05, 89.34, 89.40, 97.24, 103.65, 103.82, 113.47, 113.50, 117.72, 117.92, 127.37, 127.40, 127.98, 128.21, 128.23, 128.53, 128.64, 129.07, 130.39,

BNSDOCID: <WO_____03059926A2_I_>

130.51, 133.16, 133.67, 135.37, 135.49, 135.58, 135.70, 144.30, 144.36, 145.15, 158.91, 162.03, 171.29; ^{31}P NMR (162 MHz, $CDCl_3$) δ 150.92, 151.38. ESI-MS calculated M 919.3921, found (M + H) 920.1. Similar treatment of compound (6a) (0.23 g, 0.32 mmol) as described above for compound (6b) gave compound (7a) (0.162 g, 55.1%) as light yellow solid: R_f = 0.64, 0.81 (4.8% EtOAc in chloroform); ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.05-1.15 (m, 12H), 1.80-1.95 (m, 4H), 2.40 (m, 2H), 3.40-3.80 (m, 12H), 3.81-4.10 (m, 2H), 4.24-4.53 (m, 3H), 5.65-5.72 (m, 1H), 6.05-6.11 (m, 1H), 6.85-6.87 (m, 4H), 7.05 (br s, 1H), 7.26-7.59 (m, 13H), 7.88-7.90 (m, 2H), 8.29-8.55 (m, 1H); ^{31}P NMR (162 MHz, $CDCl_3$) δ 150.58, 151.37. ESI-MS calculated M 919.3921, found (M + H) 920.1.

B. Synthesis of an Adenosine Synthon

The next step in the chemical synthesis of 2'(3')-aminoacylated oligonucleotides is the production of an adenosine synthon. Figure 3 shows the synthesis of adenosine synthon and the coupling of the cytidine to the adenosine synthon to form 2'-*O*-THF-pCpA.

The first step in the production of the adenosine synthon is similar to that of the cytidine synthon. Generally, 6-*N*-benzoyladenosine is obtained in the same manner as 4-*N*-benzoyl-cytidine (2) and then protected with *tert*-butyldiphenylsilyl chloride at the 5'-hydroxy to form compound (9). In one embodiment, to a suspension of adenosine (10.68 g, 39.96 mmol) in pyridine (200 mL), trimethylchlorosilane (38.1 mL, 300 mmol) is added. After the reaction mixture stirred at room temperature under argon atmosphere for 45 min, benzoyl chloride (23.1 mL, 200 mmol) is added and the mixture is stirred at room temperature for 4 h. The mixture is then cooled in an ice bath and water (40 mL) is added. After 20 minutes 29% of aqueous ammonium hydroxide (96 mL, 713.3 mmol) is added slowly and the mixture is stirred at room temperature for 45 min. The reaction mixture is then evaporated to near dryness and the residue is dissolved in water (320 mL). The solution is washed once with ethyl acetate (100 mL). Crystallization begins immediately after separation of the layers. The solid is filtered out and washed by EtOAc (30 mL), ice water (40 mL), and acetone (40 mL × 2) to give 6-*N*-benzoyladenosine in 95% yield (14.17 g, 95.6%) as white solid: R_f = 0.24 (13.3% methanol in chloroform); ^{1}H NMR (400 MHz, DMSO-d_6) δ 3.55-3.70 (m, 2H), 3.96-

3.99 (m, 1H), 4.17-4.19 (m, 1H), 4.63-4.67 (m, 1H), 5.14 (t, J = 5.50 Hz, 1H), 5.26 (d, J = 4.77 Hz, 1H), 5.58 (d, J = 5.87 Hz, 1H), 6.04 (d, J = 5.51 Hz, 1H), 7.53-7.64 (m, 3H), 8.03-8.04 (m, 2H), 8.72 (s, 1H), 8.75 (s, 1H), 11.22 (br s, 1H); ^{13}C NMR (100 MHz, DMSO-d_6) δ 62.00, 71.05, 74.33, 86.40, 88.24, 126.58, 129.18, 133.17, 133.99, 143.86, 151.09, 152.33, 152.91, 166.31. 6-*N*-Benzoyladenosine is then protected with *tert*-butyldiphenylsilyl chloride at the 5'-hydroxy to form 5'-*O*-(tert-butyldiphenylsilyl)-6-*N*-benzoyladenosine (9) in 95% yield. To a solution of 6-*N*-benzoyladenosine (1.9 g, 5.12 mmol) and 4-dimethylaminopyridine (0.02 g, 0.164 mmol) in pyridine (100 mL), *tert*-butyl diphenylsilyl chloride (1.6 mL, 6.14 mmol) is injected in one portion. The mixture is stirred at room temperature under argon atmosphere for 96 h. The solvent is removed and the residue is diluted by EtOH (10 mL) and precipitated by adding Et_2O (250 mL). The solid is filtered out and washed with water (50 mL, twice) to give compound (9) (3.09 g, 99%) as white solid: R_f = 0.32 (9.1% methanol in chloroform); ^{1}H NMR (400 MHz, DMSO-d_6) δ 0.96 (s, 1H), 3.78-4.09 (m, 2H), 4.07 (m, 1H), 4.36 (m, 1H), 4.73 (t, J = 5.14 Hz, 1H), 6.06 (d, J = 5.13 Hz, 1H), 7.32-7.66 (m, 13H), 8.02-8.04 (m, 2H), 8.59 (s, 1H), 8.66 (s, 1H), 11.22 (br s, 1H); ^{13}C NMR (100 MHz, DMSO-d_6) δ 27.25, 64.62, 70.60, 73.76, 85.25, 88.46, 126.53, 128.20, 128.48, 128.53, 129.16, 129.19, 129.85, 130.59, 132.24, 133.27, 133.44, 134.00, 135.10, 135.69, 135.74, 143.78, 151.09, 152.77.

The treatment of 5'-*O*-(tert-butyldiphenylsilyl)-6-*N*-benzoyladenosine (9) with acetic anhydride in pyridine yields 2',3'-*O*,*O*-diacetyl-5'-*O*-(tert-butyldiphenylsilyl)-6-*N*-benzoyladenosine (10) in quantity yield. In one embodiment, to a solution of compound (9) (10.6 g, 17.4 mmol) and 4-dimethylaminopyridine (0.106 g, 0.87 mmol) in dry pyridine (250 mL), acetic anhydride (3.6 mL, 38.26 mmol) is added slowly by syringe. The reaction mixture is stirred at room temperature under argon atmosphere for 17.5 h, then another portion of acetic anhydride (0.164 mL, 0.132 mmol) is added. The reaction mixture is stirred for another 18 h. Water (1 mL) is added to quench the reaction. After the solvent is removed under reduced pressure, the residue was dissolved in CH_2Cl_2 (500 mL) and washed by saturated aqueous NaCl (60 mL, twice). The separated organic layer is dried over Na_2SO_4, the solvent is removed under reduced pressure, and the residue is purified by flash column chromatography (SiO_2, 4.7-40% EtOAc in hexane) to give compound (10) (23.2 g, 99%) as white solid: R_f = 0.72 (13.3% methanol in

chloroform); ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.11 (s, 9 H), 2.05 (s, 3H), 2.15 (s, 3H), 3.87 (dd, J = 11.72, 3.12 Hz, 1H), 3.99 (dd, J = 11.72, 3.12 Hz, 1H), 4.31 (m, 1H), 5.72 (dd, J = 5.47, 2.74 Hz, 1H), 5.94 (dd, J = 6.64, 5.47 Hz, 1H), 6.40 (d, J = 7.03 Hz, 1H), 7.35-7.45 (m, 6H), 7.51-7.55 (m, 2H), 7.60-7.63 (m, 1H), 7.66-7.68 (m, 4H), 8.00-8.02 (m, 2H), 8.26 (s, 1H), 8.78 (s, 1H), 8.92 (br s, 1H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 20.63, 20.92, 27.16, 63.67, 71.49, 73.72, 84.04, 84.99, 123.40, 128.06, 128.18, 128.20, 129.15, 130.25, 130.33, 132.38, 132.73, 133.08, 133.84, 135.68, 135.89, 141.11, 149.82, 152.25, 153.28, 164.74, 169.58, 169.96; ESI-MS calculated M 693.2619, found (M + H) 694.1.

The final step in the production of the adenosine synthon involves the removal of the *tert*-butyldiphenylsilyl group with tetrabutylammonium fluoride to produce 6-*N*-benzoyl-2',3'-*O,O*-diacetyladenosine (11) in 95% yield. In one embodiment, tetrabutylammonium fluoride (25.4 mL, 25.4 mmol) is injected slowly to an ice-cooled solution of 2',3'-*O,O*-diacetyl-5'-*O*-(tert-butyldiphenylsilyl)-6-*N*-benzoyladenosine (10) (17.60 g, 25.4 mmol) in tetrahydrofuran (300 mL). The reaction mixture is stirred at ice bath under argon atmosphere for 7 h. After the solvent is removed, the residue is dissolved in CH_2Cl_2 (800 mL) and washed by water (100 mL, twice) and saturated aqueous NaCl (100 mL). The organic phase was dried over Na_2SO_4 and concentrated in vacuum. The residue is purified by column chromatography (SiO_2, 0.5-2.0% MeOH in CH_2Cl_2) to give 6-*N*-benzoyl-2',3'-*O,O*-diacetyladenosine (11) in 95% yield (10.95 g, 94.8 %) as white solid: R_f = 0.23 (4.8% methanol in chloroform); ^{1}H NMR (400 MHz, $CDCl_3$) δ 2.01 (s, 3H), 2.16 (s, 3H), 3.83-4.01 (m, 2H), 4.37 (m, 1H), 5.68-5.69 (m, 1H), 5.87-5.90 (m, 1H) 6.00-6.03 (m, 1H), 6.12 (d, J = 7.6 Hz, 1H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 20.55, 20.98, 62.78, 72.85, 72.93, 86.54, 88.54, 124.82, 128.15, 129.16, 133.23, 133.53, 142.74, 150.56, 151.12, 152.64, 164.75, 169.24, 169.95; ESI-MS calculated M 455.1441, found (M - H) 454.1.

C. Chemical Coupling of Cytidine and Adenosine Synthons to Form Fully Protected pCpA Dinucleotide

Following the synthesis of the cytidine and adenosine synthons, the two can be coupled to produce a fully protected dinucleotide which is then converted to the tetrabutylammonium salt (15). The cytidine phosphoramidite (7) can be coupled with 6-

N-benzoyl-2',3'-*O,O*-diacetyladenosine (11) in the presence of 1*H*-tetrazole to yield the fully protected dinucleotide (12) in high yield. 4-*N*-benzoyl-2'-*O*-tetrahydrofuranyl-5'-*O*-(4,4'-dimethoxytrityl)cytidylyl(3' → 5'-cyanoethylphosphate)-6-*N*-Benzoyl-2',3'-*O,O*-diacetyladenosine (12) is produced in 84% yield. For compound 12: R_f= 0.50 (9.1% methanol in EtOAc); ^{1}H-NMR ($CDCl_3$) δ1.67-1.87 (m, 4H), 2.01-2.11 (m, 6H), 2.58-2.73 (m, 2H), 2.98 (br s, 1H), 3.42-3.71 (m, 4H), 3.73-3.74 (m, 6H), 4.07-4.67 (m, 6H), 4.62-4.67 (m, 1H), 5.02-5.08 (m, 1H), 5.30-5.38 (m, 1H), 5.67-5.69 (m, 1H), 5.86-5.90 (m, 1H), 6.22-6.30 (m, 2H), 6.79-6.84 (m, 4H), 7.04 (br s, 1H), 7.19-7.53 (m, 16H), 7.85-7.87 (m, 2H), 7.96-7.98 (m, 2H), 8.05-8.06 (m, 1H), 8.22-8.34 (m, 1H), 8.70-8.71 (m, 1H); ^{31}P NMR (162 MHz, $CDCl_3$) δ -1.77, -1.61; ESI-MS calcld. 1289.4, found 1328.4 $(M + K)^+$. A solution of compound (7b) (0.32 g, 0.348 mmol) in CH_3CN (5 mL) is injected to a solution of compound (11) (0.144 g, 0.316 mmol) and 1H-tetrazole (0.11 g 1.58 mmol) in CH_3CN (8 mL). The mixture is stirred at room temperature under argon atmosphere for 4 h. *Tert*-butyl hydroperoxide (0.646 mL, 3.16 mmol) is added to the mixture cooled with ice-bath and the mixture is stirred for another 2.5 h. The mixture was diluted with AcOEt (300 mL) and the organic phase is washed by saturated aqueous $NaHCO_3$ (20 mL) and saturated aqueous NaCl (20 mL). The organic layer is dried over Na_2SO_4 and the solvent is removed under reduced pressure. The residue is purified by column chromatography (SiO_2, 0.1-4% CH_3OH in CH_2Cl_2) to give compound (12b) (0.377 g, 84%,) as yellow solid: R_f= 0.50 (9.1% methanol in EtOAc); ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.67-1.87 (m, 4H), 2.01-2.11 (m, 6H), 2.58-2.73 (m, 2H), 2.98 (br s, 1H), 3.42-3.71 (m, 4H), 3.73-3.74 (m, 6H), 4.07-4.67 (m, 6H), 4.62-4.67 (m, 1H), 5.02-5.08 (m, 1H), 5.30-5.38 (m, 1H), 5.67-5.69 (m, 1H), 5.86-5.90 (m, 1H), 6.22-6.30 (m, 2H), 6.79-6.84 (m, 4H), 7.04 (br s, 1H), 7.19-7.53 (m, 16H), 7.85-7.87 (m, 2H), 7.96-7.98 (m, 2H), 8.05-8.06 (m, 1H), 8.22-8.34 (m, 1H), 8.70-8.71 (m, 1H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 19.71, 19.79, 20.60, 20.77, 20.80, 23.26, 23.34, 32.49, 32.62, 55.46, 55.49, 62.55, 62.59, 62.64, 62.84, 62.89, 66.96, 67.00, 67.27, 67.32, 67.81, 67.85, 70.42, 70.58, 73.18, 73.26, 76.31, 76.36, 76.90, 81.09, 81.16, 81.27, 82.56, 82.63, 85.97, 86.27, 87.44, 87.53, 87.68, 87.71, 97.40, 103.86, 103.90, 113.61, 116.71, 116.79, 123.95, 123.96, 127.49, 127.53, 127.98, 128.28, 128.32, 128.49, 128.53, 128.96, 129.11, 130.42, 130.46, 132.93, 133.33, 133.74, 135.09, 141.83, 141.91, 144.07, 145.04, 150.21, 151.90, 152.04, 153.01, 158.96, 158.99, 162.49, 165.24, 169.62, 169.64, 169.87, 169.97;

BNSDOCID: <WO_____03059926A2_I_>

^{31}P NMR (162 MHz, $CDCl_3$) δ -1.77, -1.61; ESI-MS calculated M 1289.4107, found (M + K) 1328.4. Similar treatment of compound (7a) (0.74 g, 0.8 mmol) to that described for (7b) and (11) (0.333 g, 0.73 mmol) gives compound (12a) (0.783 g, 83%) as light yellow solid: R_f= 0.23 (4.8% methanol in chloroform); ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.90-1.96 (m, 4H), 2.00-2.11 (m, 6H), 2.55-2.68 (m, 2H), 3.45-3.77 (m, 9H), 3.79-3.96 (m, 2H), 4.15-4.38 (m, 6H), 4.99-5.03 (m, 1H), 5.48-5.54 (m, 1H), 5.65 (m, 1H), 5.82 (m, 1H), 6.04-6.07 (m, 1H), 6.22-6.24 (m, 1H), 6.81-6.85 (m, 4H), 7.00 (br s, 1H), 7.22-7.54 (M, 16H), 7.85-7.87 (m, 2H), 7.97-7.99 (m, 2H), 8.21-8.30 (m, 3H), 8.71-8.72 (m, 1H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 19.66, 19.72, 20.58, 20.75, 23.42, 23.55, 32.70, 55.44, 55.47, 61.07, 61.37, 62.59, 62.64, 62.81, 62.85, 67.15, 67.78, 70.47, 70.56, 73.12, 73.20, 77.34, 77.74, 81.07, 81.14, 86.12, 87.58, 89.10, 89.22, 97.30, 103.72, 103.79, 113.55, 116.81, 116.89, 124.02, 124.08, 127.51, 128.06, 128.31, 128.66, 128.89, 128.97, 130.48, 132.89, 133.18, 133.34, 133.76, 135.09, 135.24, 141.99, 142.11, 143.90, 144.62, 150.22, 151.98, 152.86, 154.92, 158.95, 162.89, 165.43, 169.63, 169.82; ^{31}P NMR (162 MHz, $CDCl_3$) δ -1.41, -1.25; ESI-MS calculated M 1289.4107 found (M + Na) 1312.4, (M + K) 1328.4.

Following the production of the protected dinucleotide (12), the protecting groups can be removed. In a preferred embodiment, the DMTr group at the 5'-position of dinucleotide (12) is deprotected by hydrogenation in methanol with Pd/C (10%) catalyst. To avoid removal of the THF group at 2'-position of cytidine, due to accumulation of *in situ* formed proton in the hydrogenation process, prolonged reaction times are avoided. The deprotected product, 4-*N*-benzoyl-2'-*O*-tetrahydrofuranyl-cytidylyl(3'→5'-cyanoethylphosphate)-6-*N*-Benzoyl-2', 3'-*O*,*O*-diacetyladenosine (13) can be produced with 70% isolated yield using H_2/Pd-C in methanol and a two hour reaction time. For compound (13): R_f = 0.36 (9.1% methanol in chloroform); ^{1}H NMR ($CDCl_3$) δ1.54-1.86 (m, 4H), 2.00-2.10 (m, 6H), 2.72-2.76 (m, 2H), 3.12 (br s, 1H), 3.56-3.79 (m, 4H) 4.19-4.28 (m, 3H), 4.40-4.80 (m, 4H), 5.10-5.21 (m, 2H), 5.66-5.70 (m, 1H), 5.86-5.94 (m, 2H), 6.27-6.30 (m, 1H), 7.37-7.51 (m, 7H), 7.82-8.36 (m, 5H), 8.36 (m, 1H), 8.68 (s, 1H), 9.63 (br, 2H); ^{31}P NMR ($CDCl_3$) δ-1.69, -1.67; ESI-MS calcld. 987.2800, found 1010.2 (M + Na), 1026.1 $(M + K)^+$. More specifically, a suspension of compound (12b) (0.49 g, 0.38 mmol) and 10% Pd/C (0.10 g) in CH_3OH (20 mL) was stirred at room temperature under atmosphere hydrogen for 1.5 h. The

catalyst was filtered out and the solvent was removed. The residue was purified by flash chromatography (SiO_2, 0.1-5% CH_3OH in CH_2Cl_2) to give compound (13b) (0.263 g, 70.1%,) as yellow solid: R_f = 0.36 (9.1% methanol in chloroform); ^1H NMR (400 MHz, $CDCl_3$) δ 1.54-1.86 (m, 4H), 2.00-2.10 (m, 6H), 2.72-2.76 (m, 2H), 3.12 (br s, 1H), 3.56-3.79 (m, 4H) 4.19-4.28 (m, 3H), 4.40-4.80 (m, 4H), 5.10-5.21 (m, 2H), 5.66-5.70 (m, 1H), 5.86-5.94 (m, 2H), 6.27-6.30 (m, 1H), 7.37-7.51 (m, 7H), 7.82-8.36 (m, 5H), 8.36 (m, 1H), 8.68 (s, 1H), 9.63 (br, 2H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 19.73, 19.78, 19.81, 19.86, 20.61, 20.80, 23.35, 23.43, 32.40, 32.47, 61.43, 62.85, 62.92, 62.98, 66.87, 67.18, 67.66, 67.72, 70.45, 70.60, 73.21, 73.34, 76.86, 81.19, 81.27, 81.39, 84.65, 84.70, 85.95, 86.21, 90.88, 97.63, 104.34, 104.45, 117.03, 117.05, 123.95, 128.12, 128.37, 128.91, 129.00, 132.98, 133.30, 133.58, 133.61, 141.96, 142.01, 150.20, 152.03, 152.09, 152.93, 163.01, 165.65, 169.72, 169.73, 170.01, 170.03; ^{31}P NMR (162 MHz, $CDCl_3$) δ -1.69, -1.67; ESI-MS calculated M 987.2800, found (M + Na) 1010.2, (M + K) 1026.1. Similar treatment of compound (12a) (0.3 g, 0.232 mmol) as that described for compound (12b) and 10% Pd/C (0.1 g) gives compound (13a) (0.15 g, 65.2%) as light yellow solid: R_f = 0.45 (9.1% methanol in chloroform); ^1H NMR (400 MHz, $CDCl_3$) δ 1.79-1.95 (m, 4H), 2.04-2.11 (m, 6H), 2.26 (br s, 1H), 2.75-2.77 (m, 2H), 3.72-3.96 (m, 4H), 4.24-4.67 (m, 7H), 4.90-4.99 (m, 1H), 5.51 (m, 1H), 5.69-5.72 (m, 1H), 5.85-5.89 (m, 2H), 6.30-6.31 (m, 1H), 7.43-7.58 (m, 7H), 7.86 (m, 2H), 8.01 (m, 2H), 8.34-8.41(m, 2H), 8.75 (m, 1H), 9.32 (br, 2H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 19.80, 19.88, 20.62, 20.81, 23.52, 23.57, 32.66, 60.00, 62.91, 62.96, 67.08, 67.74, 67.83, 70.39, 70.49, 73.26, 73.36, 76.94, 81.22, 81.29, 83.16, 83.22, 85.94, 86.23, 91.09, 97.25, 104.14, 116.76, 116.98, 123.82, 127.96, 128.28, 128.34, 129.05, 129.13, 133.07, 133.35, 133.63, 141.75, 150.09, 152.01, 153.09, 162.83, 165.25, 169.72, 169.99, 170.04; ^{31}P NMR (162 MHz, $CDCl_3$) δ -1.11, -0.98; ESI-MS calculated M 987.2800, found (M + Na) 1010.2.

In another embodiment, a base (*e.g.*, pyridine, triethylamine) is added to the above deprotection reaction condition to inhibit THF cleavage. In another embodiment, ethanol is used as a solvent to replace methanol. In yet another embodiment, *tert*-butyldiphenylsilyl (TBDPS) group is used to protect the 5'-hydroxy of cytidine while the 2'-hydroxy group was protected by THF. The 2-cyanoethyl group at phosphate may not be stable under cleaving conditions for TBDPS by tetrabutylammonium fluoride.

The 5'-hydroxy-CpA is then reacted with bis(2-cyanoethoxy)(*N,N*-diisopropyl) phosphoramidite in the presence of *1H*-tetrazole in CH_3CN to yield the fully protected pCpA, 4-*N*-benzoyl-2'-*O*-tetrahydrofuranyl-5'-(di(2-cyanoethyl)-phosphate) cytidylyl (3'→5'-cyanoethylphosphate)-6-*N*-benzoyl-2',3'-*O,O*-diacetyladenosine (14). For compound (14): R_f = 0.44 (9.1% methanol in chloroform); ^{1}H NMR ($CDCl_3$) δ1.67-1.92 (m, 4H), 2.03-2.13 (m, 6H), 2.75-2.78 (m, 6H), 3.65-3.73 (m, 2H), 4.26-4.67 (m, 13H), 5.09-5.12 (m, 1H), 5.22-5.27 (m, 1H), 5.68-5.73 (m, 1H), 5.83-5.91 (m, 2H), 6.30-6.32 (m, 1H), 7.43-7.57 (m, 7H), 7.78-7.89 (m, 3H), 7.99-8.02 (m, 2H), 8.36-8.38 (m, 1H), 8.74-8.75 (m, 1H), 9.43 (br, 2H); ^{31}P NMR ($CDCl_3$) δ-1.59, -1.43, -1.41; ESI-MS calcld. 1173.2994 found 1196.1 $(M + Na)^+$, 1212.1 $(M + K)^+$. In one embodiment, compound (13b) (0.2 g, 0.203 mmol) is dried by oil pump over night and then dissolved in dry CH_3CN (30 mL). To the solution bis(2-cyanoethoxy)(*N,N*-diisopropyl)phosphoramidite (0.082 g, 0.304 mmol) and 1H-tetrazole (0.071 g, 1.01 mmol) is added. The mixture is stirred at room temperature for 3 h and tert-butyl hydroperoxide (0.2 mL, 1.01 mmol) was added. After 2 h, the solvent is removed under reduced pressure. The residue is dissolved in EtOAc (200 mL) and washed by saturated aqueous $NaHCO_3$ (20 mL) and saturated aqueous NaCl (20 mL). The organic layer is dried over Na_2SO_4 and the solvent is evaporated under reduced pressure. The residue is purified by a flash column of silica gel (0.1-5% CH_3OH in CH_2Cl_2) to give compound (14b) in 86% yield (0.204 g, 86.1%) as white solid: R_f = 0.44 (9.1% methanol in chloroform); ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.67-1.92 (m, 4H), 2.03-2.13 (m, 6H), 2.75-2.78 (m, 6H), 3.65-3.73 (m, 2H), 4.26-4.67 (m, 13H), 5.09-5.12 (m, 1H), 5.22-5.27 (m, 1H), 5.68-5.73 (m, 1H), 5.83-5.91 (m, 2H), 6.30-6.32 (m, 1H), 7.43-7.57 (m, 7H), 7.78-7.89 (m, 3H), 7.99-8.02 (m, 2H), 8.36-8.38 (m, 1H), 8.74-8.75 (m, 1H), 9.43 (br, 2H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 19.84, 19.86, 19.91, 20.61, 20.63, 20.83, 23.50, 23.56, 32.51, 32.59, 62.92, 62.97, 62.99, 66.62, 67.25, 67.86, 67.91, 70.52, 70.56, 73.18, 73.29, 74.60, 74.75, 76.45, 76.56, 81.05, 81.12, 81.24, 81.31, 85.95, 86.08, 92.40, 97.44, 105.05, 117.03, 117.14, 124.07, 124.12, 128.06, 128.33, 129.00, 129.12, 133.02, 133.42, 133.71, 141.96, 142.06, 146.60, 150.21, 152.14, 154.75, 163.04, 165.38, 169.66, 169.71, 169.98; ^{31}P NMR (162 MHz, $CDCl_3$) δ -1.59, -1.43, -1.41; ESI-MS calculated calculated M 1173.2994, found (M + Na) 1196.1, (M + K) 1212.1. Similar treatment of compound (13a) (0.096 g, 0.0972 mmol) as described for compound (13b) and bis(2-cyanoethoxy)(*N,N*-

diisopropyl)phosphoramidite (0.039 g, 0.144 mmol) gave compound (14a) (0.1g, 87.7%) as white solid: R_f= 0.44 (9.1% methanol in chloroform); ^{1}H NMR (400 MHz, $CDCl_3$) δ 1.77-1.99 (m, 4H), 2.02-2.14 (m, 6H), 2.74-2.83 (m, 6H), 3.75-3.84(m, 2H), 4.27-4.60(m, 13H), 4.85-4.90 (m, 1H), 5.54 (m, 1H), 5.68-5.73 (m, 1H), 5.80-5.88 (m, 2H), 6.29-6.33 (m, 1H), 7.44-7.58 (m, 7H), 7.88-7.92 (m, 2H), 8.00-8.05 (m, 3H), 8.38-8.41 (m, 1H), 8.72-8.76 (m, 1H), 9.37 (br, 2H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 19.82, 19.91, 19.98, 20.57, 20.63, 20.85, 23.52, 23.57, 32.14, 32.62, 63.03, 63.06, 63.08, 63.14, 65.74, 67.38, 67.43, 67.82, 67.88, 70.48, 70.58, 73.31, 76.47, 76.51, 80.17, 80.23, 81.22, 81.29, 85.81, 86.06, 90.81, 97.28, 103.99, 104.10, 116.99, 117.02, 117.06, 117.17, 123.94, 128.01, 128.31, 128.35, 129.03, 129.15, 133.04, 133.42, 133.71, 141.80, 142.12, 144.78, 150.16, 150.22, 152.09, 153.01, 154.93, 162.95, 165.26, 169.73, 169.75, 170.00, 170.02; ^{31}P NMR (162 MHz, $CDCl_3$) δ -1.34, -1.31, -1.19, -0.98 ; ESI-MS calculated M 1173.2994, found (M + Na) 1196.3.

The production of the 5'-phosphoryl-2'-*O*-tetrafuranylcytidylyl-(3'→5'-phosphoryl) adenosine tris(tetrabutylammonium) salt (15), the final step depicted in Figure 3, involves dissolving compound (14) in a mixture of ammonium hydroxide and ethanol (v/v 3/1) in a sealed reaction vessel. The mixture can be heated at 55°C overnight and then exchanged to 5'-phosphoryl-2'-*O*-tetrafuranylcytidylyl-(3'→5'-phosphoryl) adenosine tris(tetrabutylammonium) salt (15) (89%) by ion-exchange chromatography. For compound (15): R_f= 0.30 (isopropyl alcohol/ammonium/water = 7/1/2); ^{1}H NMR (D_2O) δ0.73 (m, 36H), 1.15 (m, 24H), 1.42 (m, 24H), 1.55-1.72 (m, 4H), 2.96 (m, 24H), 3.41-3.50 (m, 2H), 3.78 (m, 2H), 3.94 (m, 2H), 4.18 (m, 3H), 4.34 (m, 1H), 4.59 (m, 2H), 5.03 (m, 1H), 5.87-5.90 (m, 2H), 5.93 (d, J= 6.97 Hz, 1H), 7.86 (d, J= 7.70 Hz, 1H), 8.03 (s, 1H), 8.37 (s, 1H); ^{31}P NMR (D_2O) δ-0.46, -4.89; ESI-MS calcld. 722.1 found 721.3 (M - H). In one embodiment, compound (14a) (0.098 g, 0.0835 mmol) is dissolved in a mixture of concentrated ammonium hydroxide (9 mL) and ethyl alcohol (3 mL). The reaction mixture is stirred in a sealed pressure flask at 55 °C for 16 h. The reaction progress can be tested for completion by ^{31}P NMR. After the solvent is removed under reduced pressure, the residue in water (60 mL) is first extracted by $CHCl_3$ (15 mL, twice), then lyophilized. The solid is exchanged to tetrabutylammonium salt by using Amberlite CG-50 ion-exchange resin (100-200 mesh) in tetrabutylammonium form. The eluent can be lyophilized overnight to give 5'-

Phosphoryl-2'-*O*-tetrafuranylcytidylyl-(3'→5'-phosphoryl)adenosine tris(tetrabutylammonium) salt (15a) (0.1 g, 83.3%) as a white solid: R_f = 0.32 (isopropyl alcohol / ammonium / water, 7 / 1 / 2); ^{1}H NMR (400 MHz, D_2O) δ 0.75 (m, 36H), 1.17 (m, 24H), 1.44 (m, 24H), 1.65-1.81 (m, 4H), 2.98 (m, 24H), 3.62-3.66 (m, 1H), 3.77-4.00 (m, 5H), 4.19 (m, 2H), 4.26 (m, 1H), 4.36 (m, 1H), 4.46-4.52 (m, 2H), 5.28 (m, 1H), 5.75 (d, J = 7.42 Hz, 1H), 5.84 (d, J = 4.29 Hz, 1H), 5.90 (d, J = 6.25 Hz, 1H), 7.95 (d, J = 7.43 Hz, 1H), 8.04 (s, 1H), 8.38 (s, 1H); ^{13}C NMR (100 MHz, D_2O) δ 12.98, 19.27, 22.56, 23.22, 31.83, 58.19, 62.39, 65.05, 67.99, 70.76, 73.01, 75.29, 77.62, 84.32, 84.40, 86.89, 87.50, 96.87, 103.78, 118.71, 139.70, 141.43, 149.11, 153.00, 155.68, 157.39, 165.87; ^{31}P NMR (162 MHz, D_2O) δ 3.75, -1.03; ESI-MS calculated M 722.1462, found (M - H) 721.3. Similar treatment of compound (14b) (0.1 g, 0.0852 mmol), ammonium hydroxide (15 mL) and ethyl alcohol (5 mL) as described for compound (14a) gave compound (15b) (0.109 g, 88.6%) as white solid: R_f = 0.30 (isopropyl alcohol / ammonium / water, 7 / 1 / 2); ^{1}H NMR (400 MHz, D_2O) δ 0.73 (m, 36H), 1.15 (m, 24H), 1.42 (m, 24H), 1.55-1.72 (m, 4H), 2.96 (m, 24H), 3.41-3.50 (m, 2H), 3.78 (m, 2H), 3.94 (m, 2H), 4.18 (m, 3H), 4.34 (m, 1H), 4.59 (m, 2H), 5.03 (m, 1H), 5.87-5.90 (m, 2H), 5.93 (d, J = 6.97 Hz, 1H), 7.86 (d, J = 7.70 Hz, 1H), 8.03 (s, 1H), 8.37 (s, 1H); ^{13}C NMR (100 MHz, D_2O) δ 12.98, 19.25, 22.65, 23.20, 31.65, 58.16, 63.39, 65.24, 67.58, 70.93, 73.80, 74.64, 77.85, 83.99, 84.35, 86.46, 86.77, 97.03, 104.12, 118.65, 139.84, 141.93, 149.25, 153.00, 155.68, 157.72, 165.99; ^{31}P NMR (162 MHz, D_2O) δ -0.46, -4.89; ESI-MS calculated M 722.1462, found (M - H) 721.3.

D. Chemical Synthesis of Aminoacyl-pCpA's

Various amino acids, N-blocked amino acids, unnatural amino acids, non-amino acids, and markers can be coupled to the synthesized 2'(3')-aminoacylated oligonucleotides through an aminoacyl linkage. In one embodiment, the synthesized aminoacyl-pCpA's can be used to aminoacyl or misaminoacylate tRNA molecules. In one embodiment, an amine protecting is used to impart chemical stability to the various amino acids or derivatives thereof. Numerous amine protecting groups are known in the art (Greene, et al., Protecting Groups in Organic Synthesis 2nd Ed., John Wiley and Sons, New York, (1991)). Protecting groups that are most compatible with the integrity of the derived aminoacyl-pCpA's and aminoacyl-tRNA's include the pyroglutamyl, which is

removable by enzymatic proteolysis (Roesser, et al., S.M. *Biochemistry* (1989) 28:5185), the nitroveratryloxycarbonyl (NVOC), which is removable by photolytic cleavage (Robertson, et al., *J. Am. Chem. Soc.* (1991) 113:2722) and 4-pentenoyl groups, which are removable by mild chemical treatment such as aqueous iodine treatment (U.S. Pat. 6,245,938).

As an example, the synthesis of *N*-(4-pentenoyl)-aminoacyl-pCpAs and *N*-(4-pentenoyl)-aminoacyl-tRNAs is shown in Figure 4. In one embodiment, *N*-(4-pentenoyl)-amino acids can be produced through methods known in the literature ((e.g., Lodder, M. et al. *J. Org. Chem.* 1997, 62, 778); (Lodder, M. et al. *J. Org. Chem.* 1998, 63, 794)). In another embodiment, *N*-(4-pentenoyl)-amino acids can be produced through modified methods. Pentenoic acid can be reacted with N-hydroxysuccinimide in the presence of 4-(dimethylamino)pyridine (DMAP) and dicyclohexylcarbodiimide (DCC) to form an active ester, 4-Pentenoic acid-*N*-hydroxysuccinimide ester (17). A mixture of 4-pentenoic acid (4.1 mL, 39.95 mmol), *N*-hydroxysuccinimide (5.1 g, 43.95 mmol), DMAP (0.54 g, 4.4 mmol) and DCC (9.07 g, 43.95 mmol) are dissolved in THF (200 mL) at 0°C under nitrogen atmosphere. The reaction mixture is allowed to warm to room temperature and stirred for 36 h. The reaction mixture can be stored in the freezer overnight. After the precipitate is filtered out, the solvent is removed under reduced pressure. The residue was purified by column chromatography (0.02% CH_3OH in CH_2Cl_2) to give compound (17) (7.33 g, 93.1%) as white solid: R_f = 0.50 (3.2% methanol in chloroform); ^1H NMR (400 MHz, $CDCl_3$) δ 2.47-2.52 (m, 2H), 2.70-2.74 (m, 2H), 2.84 (m, 2H), 5.07-5.16 (m, 2H), 5.82-5.89 (m, 1H); ^{13}C NMR (100 MHz, $CDCl_3$) δ 25.81, 28.55, 30.52, 116.85, 135.38, 168.27, 169.34. For example, the reaction of compound (17) with three different amino acids in DMF gave *N*-(4-pentenoyl)-(*L*)-amino acids (18a-c). These compounds were converted to active cyanomethyl esters of *N*-(4-pentenoyl)-(*L*)-amino acids 19a-c in 60-74% yield from compound (16). The spectral data were identical with published data ((Lodder, M. et al. *J. Org. Chem.* 1997, 62, 778); (Lodder, M. et al. *J. Org. Chem.* 1998, 63, 794)). In one embodiment, to a suspension of 4-pentenoic acid-*N*-hydroxysuccinimide ester (17) (0.95 g, 4.82 mmol) and (*L*)-methionine (1.08 g, 7.23 mmol) in DMF (110 mL), diisopropylethylamine (3.4 mL, 19.29 mmol) is added. The mixture is stirred at room temperature under nitrogen atmosphere for 26 h. The precipitate is filtered out and the

filtration was condensed under reduced pressure. The residue is dissolved in CH_2Cl_2 (200 mL) and washed by H_2O (30 mL). The organic layer is dried over Na_2SO_4 and the solvent is evaporated under reduced pressure to give *N*-(4-pentenoyl)-(*L*)-methionine. The residue is dissolved in CH_3CN (40 mL) and diisopropylethylamine (3.0 mL, 17.31 mmol) and chloroacetonitrile (0.96 mL, 15.15 mmol) can be added at 0°C under nitrogen atmosphere. The mixture is allowed to reach room temperature gradually and stirred for 26 h. After the solvent is removed under reduced pressure, the residue is dissolved in CH_2Cl_2 (200 mL) and washed by H_2O (40 mL). The organic layer is dried over Na_2SO_4. After the solvent is removed, the residue is purified by column chromatography (5-50% EtOAc in hexane) to give compound *N*-(4-pentenoyl)-(*L*)-methionine cyanomethyl ester (19a) (0.78 g, 65.4%) as white solid. The spectrum data were identical with the literature. Similar treatment of compound (17) (0.324 g, 1.64 mmol) and (*L*)-phenylalanine (0.326 g, 1.97 mmol) as described for 19a can be used to produce *N*-(4-pentenoyl)-(*L*)-phenylalanine cyanomethyl ester (19b) (0.35 g, 77.5%) as white solid. The spectrum data were identical with literature. Similar treatment of compound (17) (0.6 g, 3.04 mmol) and (*S*)-leucine (0.559 g, 4.26 mmol) as that described for compound (19a) can be used to produce *N*-(4-pentenoyl)-(*L*)-leucine cyanomethyl ester (19c) (0.614 g, 80.2%) as white solid. The spectrum data were identical to the literature.

The dinucleotide pCpA can be coupled through an aminoacyl linkage to any compound. As an example, the cyanomethyl esters compounds (19a-c) can be reacted with pCpA dinucleotide (15) in DMF to yield compounds (20a-c).

In one embodiment, compounds (20a-c) are produces as follows. To a solution of compound (15) (0.05 g, 0.0296 mmol) in DMF (0.61 mL), a stock solution of compound (19a) (0.0182 g, 0.0673 mmol) in DMF (0.25 mL) is dropped slowly in several portions within 4h under nitrogen atmosphere. A mixture of NH_4Ac (50 mM, 2 mL, pH 4.5) and CH_3CN (1 mL) was added to quench the reaction and the solvent was removed under reduced pressure. The residue was redissolved in a mixture of NH_4Ac (50 mM, pH 4.5) and CH_3CN and applied to reverse phase chromatography (1-50% of CH_3CN in NH_4Ac (50 mM)). After lyophilized, the residue was applied to reverse phase chromatography (1-30% CH_3CN in 0.87 M HOAc) to give 5'-phosphorylcytidylyl(3' 5')-2'(3')-*O*-[*N*-(4-pentenoyl)-(*L*)-methionyl] adenosine (20a)

(0.01 g, 39%). R_f = 0.45 (1 - butanol / HOAc / H2O, 5 / 2 / 3); ^{1}H NMR (400 MHz, D_2O, 2:1 mixture of monoacylated diastereomers) δ 1.95 (s, 3H), 2.06-2.54 (m, 8H), 3.90 (m, 2H), 4.04 (m, 2H), 4.18 (t, J = 5.13 Hz, 1H), 4.46 (m, 1H), 4.39-4.49 (m, 3H), 4.83-4.96 (m, 3H), 5.38-5.74 (m, 3H), 6.00-6.19 (m, 2H), 7.92 (m, 1H), 8.20 (s, 1H), 8.37 (s, 1H); ^{31}P NMR (162 MHz, D_2O) δ 0.03, - 0.71; ESI-MS calculated M 865.19, found (M - H) 864.1. Similar treatment of compound (15) (0.050 g, 0.0296 mmol) and compound (19b) as that described for compound (19a) gave 5'-phosphorylcytidylyl(3' 5')-2'(3')-*O*-[*N*-(4-pentenoyl)-(*S*)-phenylalanyl] adenosine 20b (0.01 g, 38.5%) as a white solid: R_f = 0.37 (1 - butanol / HOAc / H_2O, 5 / 2 / 3); ^{1}H NMR (400 MHz, D_2O, 2:1 mixture of monoacylated diastereomers) δ; 1.98-2.21 (m, 4H), 3.05 (m, 2H), 3.91 (m, 2H), 3.97 (m, 2H), 4.17-4.23 (m, 2H), 4.69-4.85 (m, 3H), 5.27-5.58 (m, 2H), 5.69-5.74 (m, 1H), 5.84-6.11 (m, 2H), 7.00-7.26 (m, 5H), 7.90-7.92 (m, 1H), 8.14 (s, 1H), 8.33 (s, 1H); ^{31}P NMR (162 MHz, D_2O) δ 0.05, -0.76; ESI-MS calculated M 881.21, found 904.1 (M + Na). Similarl treatment of compound (15) (0.050 g, 0.0296 mmol) and compound (19c) as described for compound (20a) gave 5'-phosphorylcytidylyl(3' 5')-2'(3')-*O*-[*N*-(4-pentenoyl)-(*L*)-leucinyl] adenosine (20c) (0.012 g, 48%) as a white solid: R_f = 0.50 (1 - butanol / HOAc / H_2O, 5 / 2 / 3); ^{1}H NMR (400 MHz, D_2O, 2:1 mixture of monoacylated diastereomers) δ 0.75 (m, 6H), 1.57 (m, 3H), 2.11-2.27 (m, 4H), 3.90 (m, 2H), 4.04 (m, 2H), 4.18 (t, J = 5.14 Hz, 1H), 4.23 (m, 1H), 4.36-4.49 (m, 3H), 5.36-5.74 (m, 3H), 6.00-6.17 (m, 2H), 7.91-7.94 (m, 1H), 8.19 (s, 1H), 8.36 (s, 1H); ^{31}P NMR (162 MHz, D_2O) δ 0.03, -0.73; ESI-MS calculated M 847.23, found 870.1(M + Na).

In another embodiment, any molecule, e.g., amino acids, N-blocked amino acids, unnatural amino acids, non-amino acids, reporter groups, and markers, can be coupled to pCpA through an aminoacyl linkage through the production of cyanomethyl esters. Cyanomethyl ester production is well known to those skilled in the art. In a preferred embodiment, the method described above is utilized.

The final step in the production of aminoacyl-pCpA involves deprotection and purification. As an example, *N*-(4-pentenoyl)-aminoacyl-pCpAs (21a-c) were obtained in over 80% yield from compound (15), determined by HPLC (C18 reverse phase), following deprotection and purification. In one embodiment, THF is removed by acetifying the solution to a pH between 2 and 7, preferably to a pH between 3 and 5, and more preferably to a pH between 3 and 4.5. In one embodiment, THF is removed by

acetifying the solution with a weak acid or dilute acid. The acid can be chosen from the group consisting of acetic acid, dilute hydrochloric acid, dilute trichloracetic acid (TCA).

In a preferred embodiment, acetic acid is used to remove THF. The THF protection group of N-block-aminoacyl-dinucleotide pCpA can be stirred with 10 mM acetic acid at room temperature for 0.5 hours. The reaction mixture can be directly loaded on a C18 reverse phase column eluted with 10 mM acetic acid (pH 4.5). In another embodiment, THF protection group of N-block-aminoacyl-dinucleotide pCpA can be dissolved in 5% acetic acid and loaded on the C18 reverse phase column eluted with 5% acetic acid. The THF group of N-blocked aminoacyl pCpA can be completely deprotected under these conditions. The desired fractions can be collected and dried by lyophilizer. The product was confirmed by NMR and MS.

For compound (21a): R_f = 0.45 (n-butanol/HOAc/H2O = 5/2/3); ^{1}H NMR (D_2O) δ1.95 (s, 3H), 2.06-2.54 (m, 8H), 3.90 (m, 2H), 4.04 (m, 2H), 4.18 (t, J = 5.13 Hz, 1H), 4.46 (m, 1H), 4.39-4.49 (m, 3H), 4.83-4.96 (m, 3H), 5.38-5.74 (m, 3H), 6.00-6.19 (m, 2H), 7.92 (m, 1H), 8.20 (s, 1H), 8.37 (s, 1H); ^{31}P NMR (D_2O) δ0.03, - 0.71; ESI-MS calcld. 865.19 found 864.1 (M - H). For compound (21b): R_f = 0.37 (n-butanol/HOAc/H_2O = 5/2/3); ^{1}H NMR (D_2O) δ1.98-2.21 (m, 4H), 3.05 (m, 2H), 3.91 (m, 2H), 3.97 (m, 2H), 4.17-4.23 (m, 2H), 4.69-4.85 (m, 3H), 5.27-5.58 (m, 2H), 5.69-5.74 (m, 1H), 5.84-6.11 (m, 2H), 7.00-7.26 (m, 5H), 7.90-7.92 (m, 1H), 8.14 (s, 1H), 8.33 (s, 1H); ^{31}P NMR (D_2O) δ0.05, -0.76; ESI-MS calcld. 881.21 found 904.1 (M + Na). For compound (21c): R_f = 0.50 (n-butanol/HOAc/H_2O = 5/2/3); ^{1}H NMR (D_2O) δ0.75 (m, 6H), 1.57 (m, 3H), 2.11-2.27 (m, 4H), 3.90 (m, 2H), 4.04 (m, 2H), 4.18 (t, J = 5.14 Hz, 1H), 4.23 (m, 1H), 4.36-4.49 (m, 3H), 5.36-5.74 (m, 3H), 6.00-6.17 (m, 2H), 7.91-7.94 (m, 1H), 8.19 (s, 1H), 8.36 (s, 1H); ^{31}P NMR (D_2O) δ0.03, -0.73; ESI-MS calcld. 847.2 found 870.1(M + Na).

V. Synthesis of Aminoacyl-tRNA's

In one embodiment, an amino acyl-tRNA molecule can be formed through the ligation of a synthesized aminoacyl pCpA with a tRNA molecule lacking the two 3'-terminal nucleotides (pCpA). This chemical synthesis can be used to load any amino acid, unnatural amino acid, non-amino acid, reporter group or marker onto any tRNA

molecule lacking the 3'-terminal pCpA. The tRNA molecule can be from any species or type (e.g., initiator or suppressor tRNA). tRNA molecules that can be used in the present invention do not necessarily possess any special properties. In a preferred embodiment, the tRNA can function in a protein synthesis system. Due to the universality of the protein translation system in living systems, a large number of tRNAs can be used with both cellular and cell-free reaction mixtures. Specific tRNA molecules which recognize unique codons, such as nonsense or amber codons (UAG), are not required. However, prior to use in the present invention the tRNA must be truncated to remove the final two nucleotides. Various truncation procedures are known in the art. For example, successive treatments with periodate plus lysine, pH 8.0, and alkaline phosphatase removes 3'-terminal residues of any tRNA molecule generating tRNA-OH-3' with a mononucleotide or dinucleotide deletion from the 3'-terminus (Neu and Heppel, *J. Biol. Chem.* (1964) 239:2927-34). Alternatively, tRNA molecules may be genetically manipulated to delete specific portions of the tRNA gene. The resulting gene is transcribed producing truncated tRNA molecules (Sampson and Uhlenbeck, *Proc. Natl. Acad. Sci. USA* (1988) 85:1033-37). In one embodiment, tRNA can be partially digested with venom phosphodiesterase to obtain tRNA-C-OH lacking pCpA at its 3'-terminus. tRNAs molecules used for aminoacylation are commercially available from a number of sources and can be prepared using well-known methods from sources including *Escherichia coli*, yeast, calf liver and wheat germ cells (Sigma Chemical; St. Louis, Mo.; Promega; Madison, Wis; Boehringer Mannheim Biochemicals; Indianapolis, Ind.). Their isolation and purification mainly involves cell-lysis, phenol extraction followed by chromatography on DEAE-cellulose. Amino-acid specific tRNA, for example $tRNA^{fmet}$, can be isolated by expression from cloned genes and overexpressed in host cells and separated from total tRNA by techniques such as preparative polyacrylamide gel electrophoresis followed by band excision and elution in high yield and purity (Seong et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:334-338). Run-off transcription allows for the production of any specific tRNA in high purity, but its applications can be limited due to lack of post-transcriptional modifications (Bruce et al., *Biochemistry* (1982) 21:3921).

As an example, *N*-(4-pentenoyl)-aminoacyl-tRNAs (22a-c) were prepared by T4 RNA ligase-mediated ligation with the chemically synthesized aminoacyl pCpAs (21a-c) and *Escherichia coli* $tRNA(-CA)^{Phe}$ (23) transcripts. In one embodiment, the ligation

reaction comprises *E. coli* tRNA(-CA)^{Phe} 23 can be performed by *in vitro* transcription in the presence of T7 RNA polymerase. In a typical reaction, 40 μg of L-aminoacyl-pCpA (22a-c or 24) can be incubated with tRNA(-CA)^{Phe} in the presence of 200 U T4 RNA ligase in the ligation buffer [55 mM HEPES (pH 7.5), 250 μM ATP, 15 mM $MgCl_2$. 10% (v/v) DMSO] at 37 °C for 25 min.-5 hours. The 4-pentenoyl protection group can be removed under 10 mM I_2 for 20 min (Noren, C. J. et al. *Nucleic Acids Res.* (1990) 18, 83); (Bhuta, A et al. *Biochemistry* (1981) 20, 8-15); (Moris et al. *Tetrahedron* (1994) 50, 6927); (Mercer et al. *Eur. J. Biochem.* (1972) 28, 38. (Kumar, G. et al. *J. Org. Chem.* (1982) 47, 634); (Happ et al. *J. Org. Chem.* (1987) 52, 5387).

A. Use of Synthesized aminoacyl-tRNA's

The aminoacyl-tRNA molecules can be placed in a translation system comprising a ribosome and messenger RNA (mRNA) under conditions suitable for a peptidyl transferase reaction, thereby synthesizing a peptide or protein incorporating the amino acid, non-amino acids, unnatural amino acid, reporter groups, and/or marker. Translation systems may be cellular or cell-free, and may be prokaryotic or eukaryotic. In one embodiment, the translation system comprises a cell-free translation system. Cell-free translation systems are commercially available and many different types and systems are well-known. Examples of cell-free systems include prokaryotic lysates such as *Escherichia coli* lysates, and eukaryotic lysates such as wheat germ extracts, insect cell lysates, rabbit reticulocyte lysates, frog oocyte lysates and human cell lysates. Eukaryotic extracts or lysates may be preferred when the resulting protein is glycosylated, phosphorylated or otherwise modified because many such modifications are only possible in eukaryotic systems. Some of these extracts and lysates are available commercially (Promega; Madison, Wis.; Stratagene; La Jolla, Calif.; Amersham; Arlington Heights, Ill.; GIBCO/BRL; Grand Island, N.Y.). Membranous extracts, such as the canine pancreatic extracts containing microsomal membranes, are also available which are useful for translating secretory proteins. Mixtures of purified translation factors have also been used successfully to translate mRNA into protein as well as combinations of lysates or lysates supplemented with purified translation factors such as initiation factor-1 (IF-1), IF-2, IF-3 (alpha or beta), elongation factor T (EF-Tu), or termination factors. Cell-free systems may also be coupled transcription/translation

systems wherein DNA is introduced to the system, transcribed into mRNA and the mRNA translated as described in Current Protocols in Molecular Biology (Ausubel et al., Wiley Interscience, 1993), which is hereby specifically incorporated by reference. RNA transcribed in eukaryotic transcription system may be in the form of heteronuclear RNA (hnRNA) or 5'-end caps (7-methyl guanosine) and 3'-end poly A tailed mature mRNA, which can be an advantage in certain translation systems. For example, capped mRNAs are translated with high efficiency in the reticulocyte lysate system.

In another embodiment, the translation system is a cellular translation system selected from the group consisting of tissue culture cells, primary cells, cells *in vivo*, isolated immortalized cells, human cells and combinations thereof. Cellular translation systems include whole cell preparations such as permeabilized cells or cell cultures wherein a desired nucleic acid sequence can be transcribed to mRNA and the mRNA translated.

In one embodiment, the chemically misaminoacylated tRNAs can be introduced into a cellular or cell-free protein synthesizing system, the translation system, where they function in protein synthesis to incorporate markers or non-native amino acids in the growing peptide chain. The translation system comprises macromolecules including RNA and enzymes, translation, initiation and elongation factors, and chemical reagents. RNA of the system is required in three molecular forms, ribosomal RNA (rRNA), messenger RNA (mRNA) and transfer RNA (tRNA). mRNA carries the genetic instructions for building a peptide encoded within its codon sequence. tRNAs contain specific anti-codons which decode the mRNA and individually carry amino acids into position along the growing peptide chain. Ribosomes, complexes of rRNA and protein, provide a dynamic structural framework on which the translation process, including translocation, can proceed. Within the cell, individualized aminoacyl tRNA synthetases bind specific amino acids to tRNA molecules carrying the matching anti-codon creating aminoacylated or charged tRNAs by the process of aminoacylation. The process of translation including the aminoacylation or charging of a tRNA molecule is described in Molecular Cell Biology (J. Darnell et al. editors, Scientific American Books, N.Y., N.Y. 1991), which is hereby specifically incorporated by reference. Aminoacylation may be natural or by artificial means utilizing native amino acids, non-native amino acid, amino acid analogs or derivatives, or other molecules such as detectable chemicals or coupling

agents. The resulting misaminoacylated tRNA comprises a native amino acid coupled with a chemical moiety, non-native amino acid, amino acid derivative or analog, or other detectable chemicals. These misaminoacylated tRNAs can incorporate their markers into the growing peptide chain during translation forming labeled nascent proteins which can be detected and isolated by the presence or absence of the marker.

Translations in cell-free systems generally require incubation of the ingredients for a period of time. Incubation times range from about 5 minutes to many hours, but is preferably between about thirty minutes to about five hours and more preferably between about one to about three hours. Incubation may also be performed in a continuous manner whereby reagents are flowed into the system and nascent proteins removed or left to accumulate using a continuous flow system (A. S. Spirin et al., *Sci.* (1988) 242:1162-64). This process may be desirable for large scale production of nascent proteins. Incubations may also be performed using a dialysis system where consumable reagents are available for the translation system in an outer reservoir which is separated from larger components of the translation system by a dialysis membrane (Kim et al. (1996) *Biotechnol Prog* 12, 645-649). Incubation times vary significantly with the volume of the translation mix and the temperature of the incubation. Incubation temperatures can be between about 4°C. to about 60°C., and are preferably between about 15°C. to about 50°C., and more preferably between about 25°C. to about 45°C. and even more preferably at about 25° C. or about 37°C. Certain markers may be sensitive to temperature fluctuations and in such cases, it is preferable to conduct those incubations in the non-sensitive ranges. Translation mixes will typically comprise buffers such as Tris-HCl, Hepes or another suitable buffering agent to maintain the pH of the solution between about 6 to 8, and preferably at about 7. Other reagents which may be in the translation system include dithiothreitol (DTT) or 2-mercaptoethanol as reducing agents, RNasin to inhibit RNA breakdown, and nucleoside triphosphates or creatine phosphate and creatine kinase to provide chemical energy for the translation process.

In cellular protein synthesis, the chemically aminoacylated or misaminoacylated tRNAs can be introduced into intact cells, cell organelles, cell envelopes and other discrete volumes bound by an intact biological membrane, which contain a protein synthesizing system. This can be accomplished through a variety of methods that have

been previously established such as sealing the tRNA solution into liposomes or vesicles which have the characteristic that they can be induced to fuse with cells. The fusion of cells is used to refer to the introduction of the liposome or vesicle interior solution containing the tRNA into the cell. Alternatively, some cells will actively incorporate liposomes into their interior cytoplasm through phagocytosis. The tRNA solution could also be introduced through the process of cationic detergent mediated lipofection (Felgner et al., *Proc. Natl. Acad. Sci. USA* (1987) 84:7413-17), or injected into large cells such as oocytes. Injection may be through direct perfusion with micropipettes or through the method of electroporation.

Alternatively, cells can be permeabilized by incubation for a short period of time in a solution containing low concentrations of detergents in a hypotonic media. Useful detergents include Nonidet-P 40 (NP40), Triton X-100 (TX-100) or deoxycholate at concentrations of about 0.01 nM to 1.0 mM, preferably between about 0.1 μM to about 0.01 mM, and more preferably about 1 μM. Permeabilized cells allow reporter groups to pass through cellular membranes unaltered and be incorporated into nascent proteins by host cell enzymes. Such systems can be formed from intact cells in culture such as bacterial cells, primary cells, immortalized cell lines, human cells or mixed cell populations. These cells may, for example, be transfected with an appropriate vector containing the gene of interest, under the control of a strong and possibly regulated promoter. Messages are expressed from these vectors and subsequently translated within cells. Intact misaminoacylated tRNA molecules, charged with a non-radioactive marker could be introduced to cells and incorporated into translated product.

VI. Uses

Improved chemical synthesis of 2'(3')-O-aminoacyl-oligonucleotides, 2'(3')-O-aminoacyl-pCpA's and aminoacyl-tRNA's can be used to more efficiently explore the structure of the ribosome as well as to introduce unnatural amino acids or reporter groups into proteins. The improved synthesis disclosed herein can be used to efficiently label proteins synthesized in a translation system using methods known to one skilled in the art. There are numerous uses and applications of chemically synthesized aminoacyl-oligonucleotides and misaminoacylated tRNA's as described by Dougherty DA. (*Curr Opin Chem Biol* (2000) 4(6):645-652), Gillmore et al. (*Topics Curr Chem* (1999) 202:

77-99), U.S. Pat. No 5,556,768 and U.S. Pat. No. 6,306,628, which are incorporated by reference. For example, proteins can be labeled while being synthesized with detectable markers which are incorporated into the peptide chain. Non-cognate amino acids, unnatural amino acids and reporter groups can be introduced into nascent proteins from the chemically synthesized misaminoacylated tRNAs during the translation process.

Any proteins that can be expressed by translation in a cellular or cell-free translation system may be nascent proteins and consequently, labeled, detected and isolated by the methods of the invention. Examples of such proteins include enzymes such as proteolytic proteins, cytokines, hormones, immunogenic proteins, carbohydrate or lipid binding proteins, nucleic acid binding proteins, human proteins, viral proteins, bacterial proteins, parasitic proteins and fragments and combinations. Products of recombinant genes and gene fusion products can be produced using this method since recombinant vectors carrying such genes typically carry strong promoters that transcribe mRNAs at fairly high levels. These mRNAs are easily translated in a translation system.

In one embodiment, the functional properties of the nascent protein can be preserved. A subset of tRNAs which will incorporate amino acid derivatives at sites which do not interfere with protein function or structure can be chosen. Amino acids at the amino or carboxyl terminus of a polypeptide do not alter significantly the function or structure. tRNA molecules which recognize the universal codon for the initiation of protein translation (AUG), when misaminoacylated, will place the amino acid derivative at the amino terminus. Prokaryotic protein synthesizing systems utilize initiator $tRNA^{fMet}$ molecules and eukaryotic systems initiator $tRNA^{Met}$ molecules. In either system, the initiator tRNA molecules are aminoacylated amino acid derivatives which may be unnatural amino acids or amino acid analogs or derivatives that possess marker, reporter or affinity properties. The resulting nascent proteins will be exclusively labeled at their amino terminus, although markers placed internally do not necessarily destroy structural or functional aspects of a protein. For example, a $tRNA^{LYS}$ may be misaminoacylated with the amino acid derivative dansyllysine which does not interfere with protein function or structure. In addition, using limiting amounts of misaminoacylated tRNAs, it is possible to detect and isolate nascent proteins having only a very small fraction labeled with marker which can be very useful for isolating

proteins when the effects of large quantities of marker would be detrimental or are unknown.

Chemically aminoacylated or misaminoacylated tRNAs can be introduced into a cellular or cell-free protein synthesis system. In the cell-free protein synthesis system, the reaction mixture contains all the cellular components necessary to support protein synthesis including ribosomes, tRNA, rRNA, spermidine and physiological ions such as magnesium and potassium at appropriate concentrations and an appropriate pH. Reaction mixtures can be normally derived from a number of different sources including wheat germ, *E. coli* (S-30), red blood cells (reticulocyte lysate,) and oocytes, and once created can be stored as aliquots at about +4°C to -70°C. The method of preparing such reaction mixtures is described by J. M. Pratt (Transcription and Translation, B. D. Hames and S. J. Higgins, Editors, p. 209, IRL Press, Oxford, 1984) which is hereby incorporated by reference. Many different translation systems are commercially available from a number of manufacturers.

The chemically aminoacylated or misaminoacylated tRNA can be added directly to the reaction mixture at a solution of predetermined volume and concentration as determined by one skilled in the art. This can be done directly prior to storing the reaction mixture at -70°C. in which case the entire mixture is thawed prior to initiation of protein synthesis. Efficient incorporation of amino acids, non-native amino acids, unnatural amino acids, reporter groups and markers into nascent proteins is sensitive to the final pH and magnesium ion concentration. Reaction mixtures are normally about pH 6.8 and contain a magnesium ion concentration of about 3 mM. These conditions impart stability to the base-labile aminoacyl linkage of the misaminoacylated tRNA. Aminoacylated tRNAs can be produced in sufficient quantities from the translation extract. Misaminoacylated tRNAs can be added at between about 1.0 μg/ml to about 1.0 mg/ml, preferably at between about 10 μg/ml to about 500 μg/ml, and more preferably at about 150 μg/ml.

Initiation of protein synthesis occurs upon addition of a quantity of mRNA or DNA to the reaction mixture containing the misaminoacylated tRNAs. mRNA molecules may be prepared or obtained from recombinant sources, or purified from other cells by procedure such as poly-dT chromatography. One method of assuring that the proper ratio of the reaction mixture components is to use predetermined volumes that

are stored in convenient containers such as vials or standard microcentrifuge tubes. For example, DNA and/or mRNA coding for the nascent proteins and the misaminoacylated tRNA solution are premixed in proper amounts and stored separately in tubes. Tubes are mixed when needed to initiate protein synthesis.

In one embodiment, the misaminoacylation can be used to detect a nascent protein. A tRNA molecule is misaminoacylated with a marker which is highly fluorescent when excited with UV (ultraviolet) radiation. The misaminoacylated tRNA is then introduced into a cell-free protein synthesis extract and the nascent proteins containing the marker analog produced. Proteins in the cell-free extract are separated by polyacrylamide gel electrophoresis (PAGE). The resulting gel contains bands which correspond to all of the proteins present in the cell-free extract. The nascent protein is identified upon UV illumination of the gel by detection of fluorescence from the band corresponding to proteins containing marker. Detection can be through visible observation or by other conventional means of fluorescence detection.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references, patents and published patent applications cited throughout this application, are incorporated herein by reference.

Examples

Example 1: Materials and Methods

Anhydrous solvents were dried and freshly distilled (THF from sodium/benzophenone, pyridine, CH_2Cl_2, CH_3CN from CaH_2), anhydrous DMF was purchased from Aldrich. ^1H, ^{13}C and ^{31}P NMR spectra were recorded at 400, 100, and 162 MHz. ^1H NMR assignments were supported by COSY experiments, while ^{13}C NMR assignments were supported by hetero COSY (HMQC) experiments. Mass spectra were recorded in ThermoFinnigan LCQ Duo. TLC was performed on EM Sciences precoated silica gel F-254 plates and viewed by UV. Chromatography refers to flash column chromatography. *Escherichia coli* tRNA(-CA)^{Phe} was prepared with T7 RNA polymerase by run-off transcription from *Escherichia coli* cDNA template.

Example 2: Ligation reaction of chemically synthesized aminoacyl-pCpA with tRNA(-CA) to form *N*-biotinylamidocaproyl-*L*-methionyl tRNA (tRNA-*O*-Met-biotin) (25)

This study demonstrates the successful formation of tRNA-*O*-Met-biotin (25) utilizing the methods of the present invention. In order to determine the ligation efficiency, *N*-biotinylamidocaproyl-*L*-methionyl-pCpA (24) and *N*-biotinylamidocaproyl-*L*-methionyl-tRNA were also synthesized in the same manner as described above.

To a solution of *N*-biotinylamidocaproyl-*L*-methione (0.958 g, 1.96 mmol) in DMF (20 mL), diisopropylethylamine (2.4 mL, 13.72 mmol) and chloroacetonitrile (0.24 mL, 3.92 mmol) were added at 0°C under nitrogen atmosphere. The reaction was allowed to stir at room temperature over night. The solvent was removed and the residue was washed by acetone (6 mL, twice). The solid was dried in vacuum over night, and then the solid was suspend in CH_2Cl_2 (15 mL) for 1 h. The solid was filtered out and was washed by CH_2Cl_2 (10 mL, twice) and ethyl ether (10 mL, twice), then the solid was dried in vacuum over night to give *N*-biotinylamidocaproyl-(*L*)-methionyl-cyanomethyl ester (0.75 g, 72.8%) as white solid. ^{1}H NMR (400 MHz, DMSO-d_6) δ 1.52-1.64 (m, 12H), 1.87-2.10 (m, 9H), 2.41-2.6(m, 3H), 2.81 (dd, J = 12.48, 5.14 Hz, 1H,), 2.99 (m, 2H), 3.08 (m, 1H), 4.11 (m, 1H), 4.28 (m, 1H), 4.40 (m, 1H), 4.99 (s, 2H), 6.37 (s, 1H), 6.43 (s, 1H), 7.74 (t, J = 5.13 Hz, 1H), 8.36 (d, J = 7.33 Hz, 1H); ^{31}C NMR (400 MHz, DMSO-d_6) δ 15.18, 25.57, 26.00, 26.66, 28.71, 28.91, 29.63, 30.03, 30.72, 35.47, 35.88, 38.96, 40.53, 50.11, 51.30, 56.12, 59.85, 61.71, 116.50, 163.38, 171.84, 172.46, 173.37.

The coupling reaction of *N*-biotinylamidocaproyl-*L*-methionyl-cyanomethylester with the tris(tetrabutylammonium) salt of pCpA (15) in anhydrous DMF yielded *N*-biotinylamidocaproyl-*L*-methionyl-pCpA (24). To a solution of pCpA (15) (0.035 g, 0.0242 mmol) in DMF (0.34 mL), *N*-biotinylamidocaproyl-(*L*)-methionyl-cyanomethyl ester was added under nitrogen atmosphere. The mixture was stirred for 2 h. A mixture of NH_4Ac (50 mM, 2 mL, pH 4.5) and CH_3CN (1 mL) was added to quench the reaction and the solvent was removed. The residue was applied to reverse phase chromatography (1-50% of CH_3CN in NH_4Ac (50 mM)). After lyophilized dry, the residue was applied to reverse phase chromatography (1-50% of CH_3CN in acetic acid (0.87 M)) to give

compound (24) (0.005 g, 21.4% for two steps) as white solid after the lyophilized: R_f = 0.32 (1-butylalcohol / acetic acid / water, 5 / 2 / 3); ^{1}H NMR (400 MHz, D_2O, 2:1 mixture of monoacylated diastereomers) δ 1.12-1.48 (m, 12H), 1.91-2.16 (m, 9H), 2.32-2.55 (m, 3H), 2.72-2.77 (m, 1H), 2.93-2.98 (m, 2H), 3.04-3.09 (m, 1H), 3.85-4.86 (m, 12H), 5.35-5.53 (m, 1H), 5.67-5.72 (m, 1H), 5.96-6.15 (m, 2H), 7.90-7.93 (m, 1H), 8.19-9.21 (m, 1H), 8.35-8.42 (m, 1H); ESI-MS calculated (M - H) 1121.3, found 1121.3.

The biotinylated aminoacyl-pCpA was ligated with an *Escherichia coli* tRNA(-CA)^{Phe} (23) transcript in the presence of T4 RNA ligase. The resulting *N*-biotinylamidocaproyl-*L*-methionyl tRNA (25) and ligation efficiency were determined by streptavidin gel-shift analysis shown in Figure 5. The samples were run on 7.5 M urea/10% polyacrylamide gel with 1 x TBE buffer at 30 W. Over 90% of tRNA(-CA) was converted to 25 within 25 min (lane 5). The result demonstrated that the ligation reaction of aminoacyl-pCpA with tRNA(-CA) was highly efficient and can be used to charge tRNA with any amino acid, unnatural amino acid, labeled amino acid, reporter group, marker or derivative thereof.

Example 3: Peptidyltransferase Activity of tRNA-*O*-Met-Biotin (25)

The successful use of the chemically synthesized tRNA-*O*-Met-Biotin (25) substrate, formed using the methods of the present invention, for the peptidyl transferase reaction in the ribosome and 30S extract is demonstrated in this study. This reaction is depicted in Figure 6. The formed peptide can be separated from the reactant and observed on 24%/7.5 M urea polyacrylamide gel when ^{32}P-CCA-*NH*-Phe as a peptidyl acceptor. Synthesis of CCA--*NH*-Phe can be performed as described by Zhang and coworkers (Zhang et al. *Org Lett* (2002) 4 (21): 3615-3618). Figure 7 is a typical gel showing the peptidyl transferase activity with the chemically synthesized tRNA-*O*-Met-Biotin (25). Reactions were performed in 70S ribosome with 50 μM tRNA-Met-Biotin (25), trace amount of ^{32}P-CCA-*NH*-Phe (~ 5×10^{-4} μM) in the presence of 50 mM Tris•HCl (pH 7.5), 5 mM $MgCl_2$, 10 mM NH_4Cl, 100 mM KCl and 1.5 mM spermidine at 37 ^{0}C. The lower band corresponds to labeled ^{32}P-CCA-*NH*-Phe; the middle band is the formed dipeptide (^{32}P-CCA-*NH*-Phe-Met-Biotin) labeled by biotin; the highest band represents ^{32}P-CCA-*NH*-Phe-Met-Biotin::streptavidin complex. When ^{32}P-CCA-*NH*-Phe was incubated with tRNA-Met-Biotin without ribosome, no product was observed

BNSDOCID. <WO_____03059926A2_I_>

(Figure 7, lanes 1-3). When tRNA-Met-Biotin (25) was incubated with ^{32}P-CCA-*NH*-Phe in the presence of 70S ribosome, a new band was formed (Figure 7, lanes 4-7). Upon streptavidin incubation, the product was shifted to the top of the gel (Figure 7, lane 7). This suggested that a peptide bond was formed between pCpCpA-*NH*-Phe and tRNA-*O*-Met-biotin. These results demonstrated that tRNA-*O*-Met-biotin (25) is fully active for the peptidyl transferase reaction in the ribosome. Therefore, the methods of this invention for the chemical synthesis of aminoacyl-tRNA's (e.g., tRNA-*O*-Met-biotin (25)) is a useful technology for studying the peptide bond formation in the ribosome and protein synthesis in the ribosome. This method can provide high yields of any aminoacyl-charged tRNA, which can be used as substrates in the ribosomal reaction or protein synthesis.

It should be evident to one of ordinary skill in the art to recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims. All references cited herein are incorporated by reference in their entirety.

What is claimed is:

1. A method of synthesizing 2'(3')-O-aminoacyl-oligonucleotides comprising:

selecting a first nucleotide;

protecting a 2'- hydroxyl group of the first nucleotide with a tetrahydrofuranyl (THF) group;

protecting a 5'-hydroxyl of the first nucleotide with a protecting group, such that the 5'-hydroxyl protecting group can be selectively removed;

selecting a second nucleotide;

coupling the first protected nucleotide with the second nucleotide, forming a protected dinucleotide;

selectively deprotecting the 5'-hydroxyl of the dinucleotide;

reacting the dinucleotide with an aminoacyl-cyanomethyl ester, forming a THF-protected 2'(3')-O-aminoacyl-oligonucleotide; and

removing the THF group, whereby a 2'(3')-O-aminoacyl-oligonucleotide is formed.

2. The method of claim 1, wherein the method further comprises selecting the nucleotides from the group consisting of guanidine, adenosine, thymidine, cytidine, and uridine.

3. The method of claim 1, wherein the method further comprises adding a base, such that THF is initially protected from cleavage.

4. The method of claim 1, wherein the method further comprises selecting the protecting group from the group consisting of dimethoxytrityl (DMTr), levulinoyl, and 4, 4', 4"- tris (4,5-dichlorophthalimido) trityl groups, tert-butyldiphenylsilyl (TBDPS), and silyl ether groups.

5. The method of claim 4, wherein the step of selecting a protecting group further comprises selecting dimethoxytrityl (DMTr).

6. The method of claim 1, wherein the step of selectively deprotecting the 5'-hydroxyl further comprises hydrogenating in alcohol.

7. The method of claim 1, wherein the method further comprises forming an aminoacyl-cyanomethylester from a member of the group consisting of natural amino acids, unnatural amino acids, amino acid derivatives, reporter groups, protected amino acids, N-blocked amino acids, N-(4-pentenoyl)-(L)-amino acids and derivatives thereof.

8. The method of claim 1, wherein the step of removing THF further comprises decreasing the pH.

9. The method of claim 8, wherein the step of removing THF further comprises adding a mild acid selected from the group comprising acetic acid, dilute hydrochloric acid, dilute trichloracetic acid (TCA).

10. The method of claim 2, wherein the method further comprises selecting cytidine as the first nucleotide and selecting adenosine as the second nucleotide.

11. The method of claim 10, wherein the method further comprises forming 2'(3')-O-aminoacyl-pCpA.

12. The method of claim 11, wherein the method further comprises a step of ligating the 2'(3')-O-aminoacyl-pCpA with a tRNA molecule lacking a CA sequence at a 3'-end using a RNA ligase, thereby producing an aminoacyl-tRNA molecule.

13. The method of claim 12, wherein the step of ligating further comprises utilizing T4 RNA ligase.

14. The method of claim 12, wherein the method further comprises selecting the tRNA molecule lacking a 3'-CA sequence.

15. The method of claim 14, wherein the step of selecting a tRNA molecule further comprises selecting an initiator tRNA molecule.

16. The method of claim 14, wherein the step of selecting a tRNA molecule further comprises selecting a suppressor tRNA molecule.

17. The method of claim 12, wherein the method further comprises incubating the aminoacyl-tRNA in a translation system suitable for a peptidyl transferase reaction, thereby synthesizing a peptide or protein.

18. The method of claim 17, wherein the method further comprises selecting a cell-free translation system.

19. The method of claim 17, wherein the method further comprises selecting a cellular translation system.

20. The method of claim 17, wherein the method further comprises synthesizing a protein comprising natural amino acids, unnatural amino acids, amino acid derivatives, labeled amino acid derivatives, reporter groups, and combinations thereof.

Figure 1

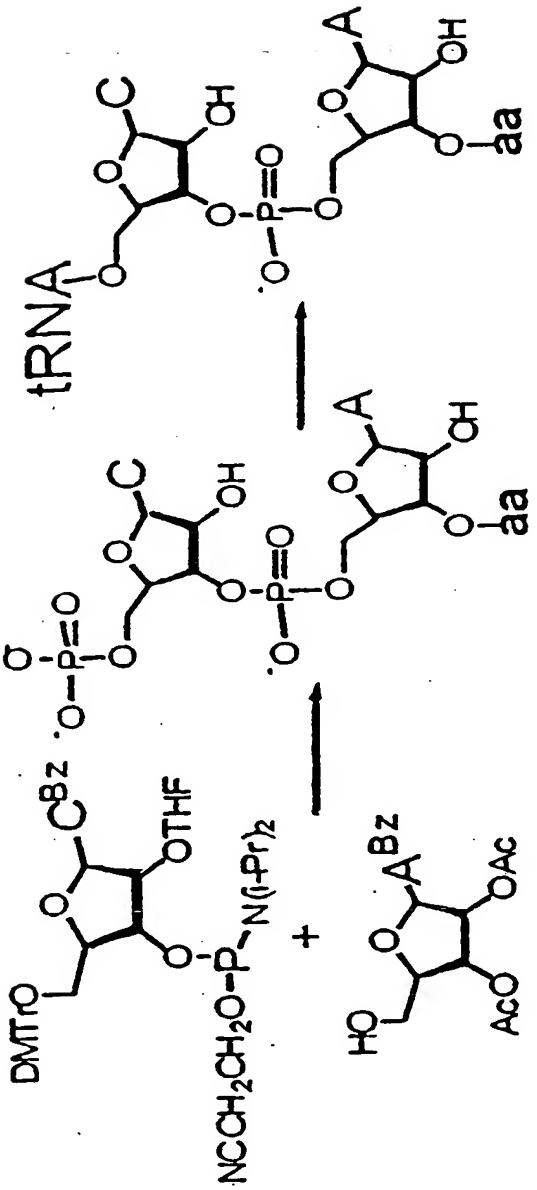


Figure 2

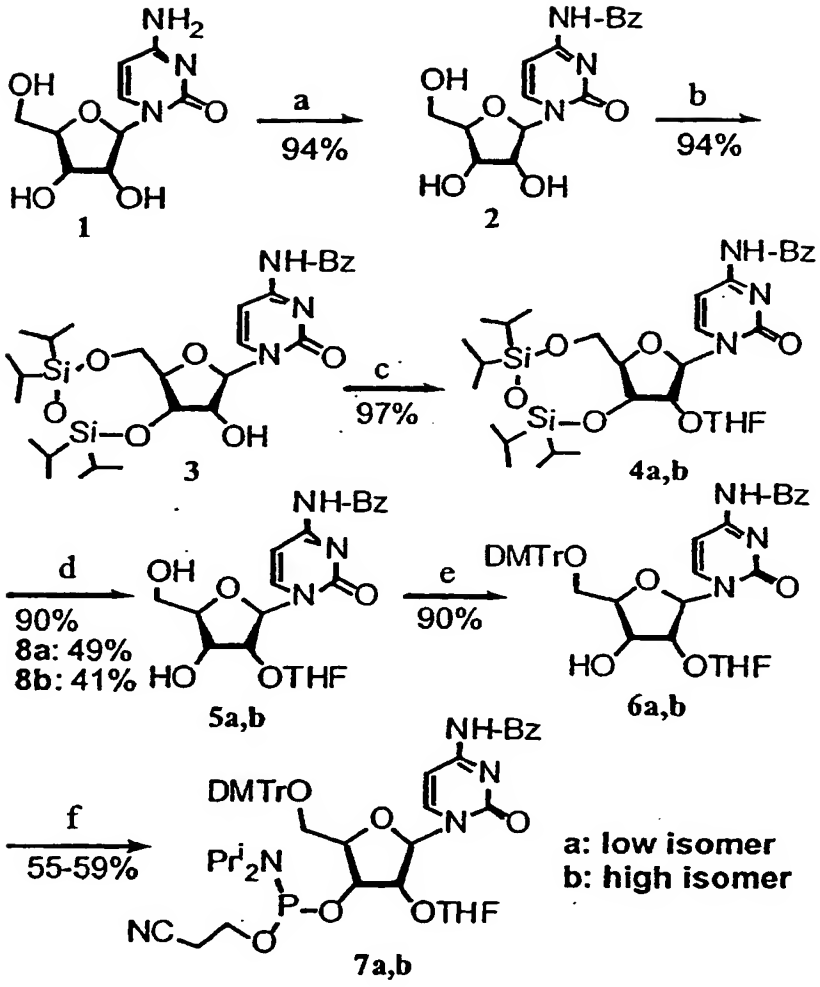

NH2
OH
HO
OH
1
a
94%
NH-Bz
2
b
94%
3
c
97%
4a,b
d
90%
8a: 49%
8b: 41%
5a,b
e
90%
DMTrO
6a,b
f
55-59%
Pr$^i_2$N
NC
7a,b
a: low isomer
b: high isomer

Figure 3

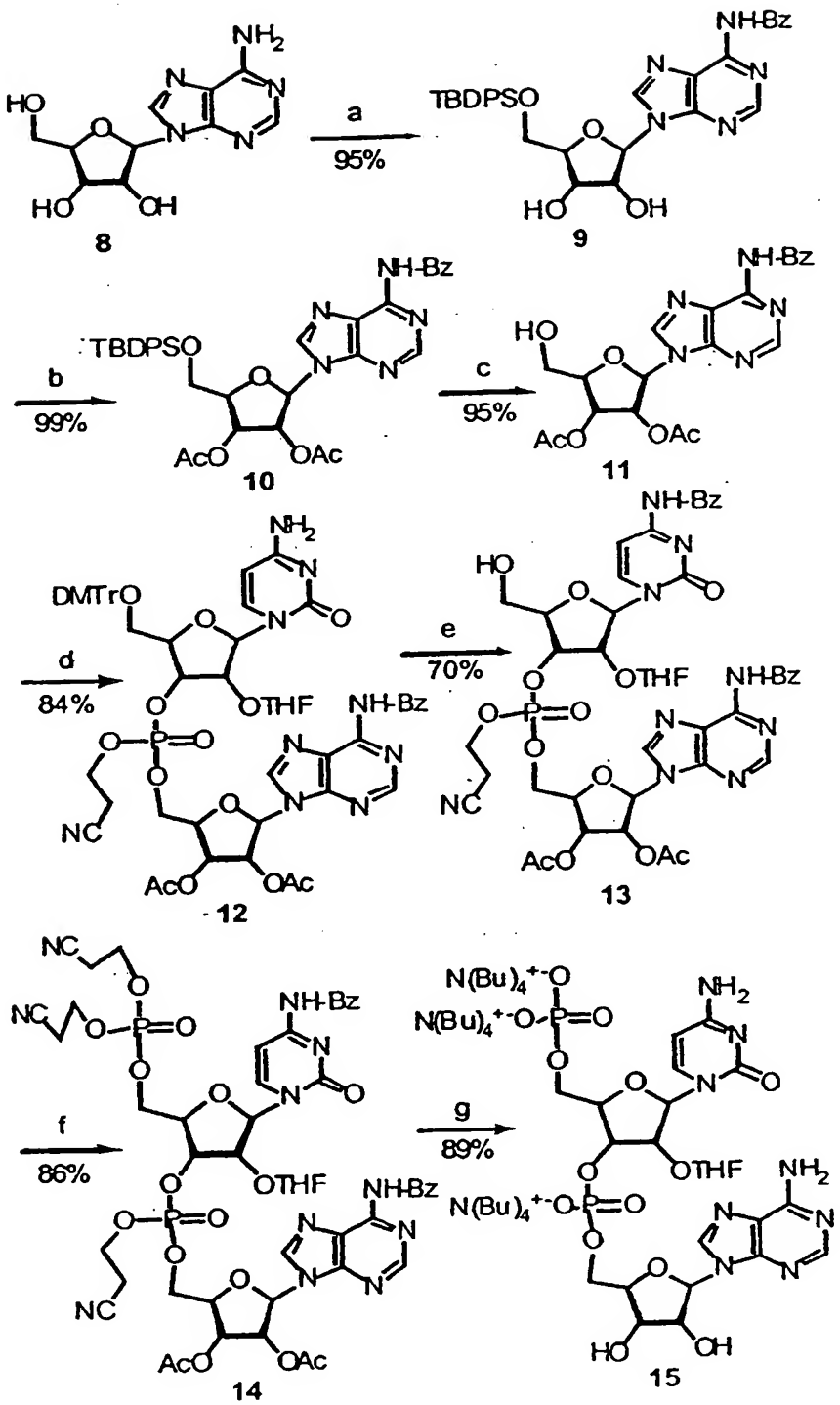
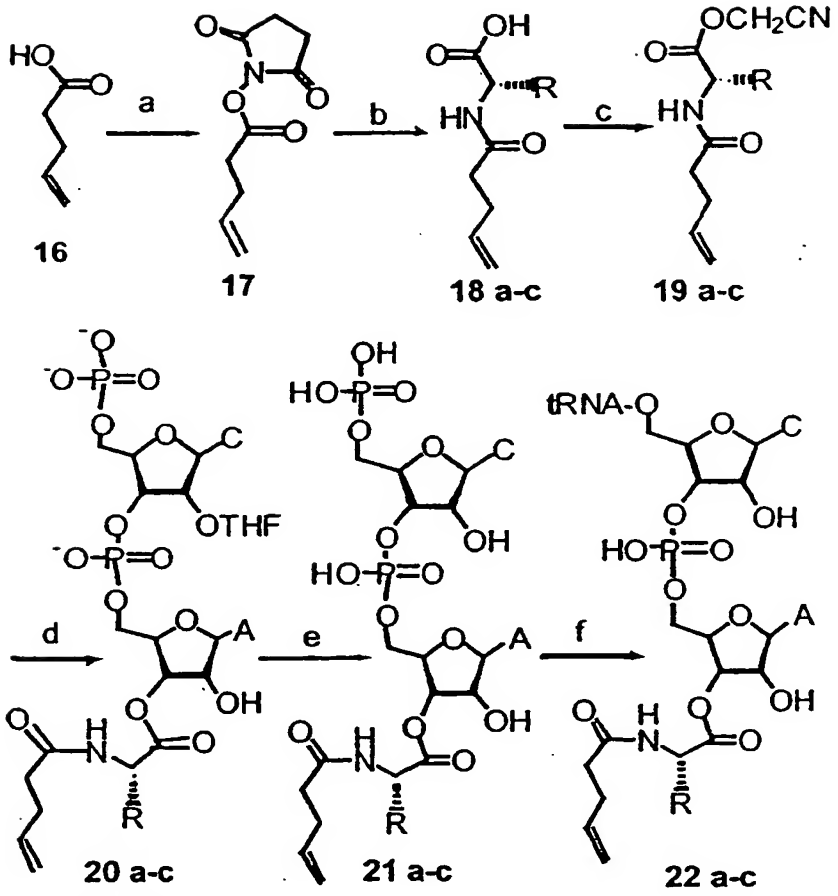

NH2
HO
HO
OH
8
a
95%
NH-Bz
TBDPSO
HO
OH
9
b
99%
NH-Bz
TBDPSO
AcO
OAc
10
c
95%
NH-Bz
HO
AcO
OAc
11
d
84%
NH2
DMTrO
OTHF
NH-Bz
NC
AcO
OAc
12
e
70%
NH-Bz
HO
OTHF
NH-Bz
NC
AcO
OAc
13
f
86%
NC
NC
NH-Bz
OTHF
NH-Bz
NC
AcO
OAc
14
g
89%
N(Bu)4+-O
N(Bu)4+-O
NH2
OTHF
NH2
N(Bu)4+-O
HO
OH
15

Figure 4



BNSDOCID: <WO_____03059926A2_I_>

Figure 5

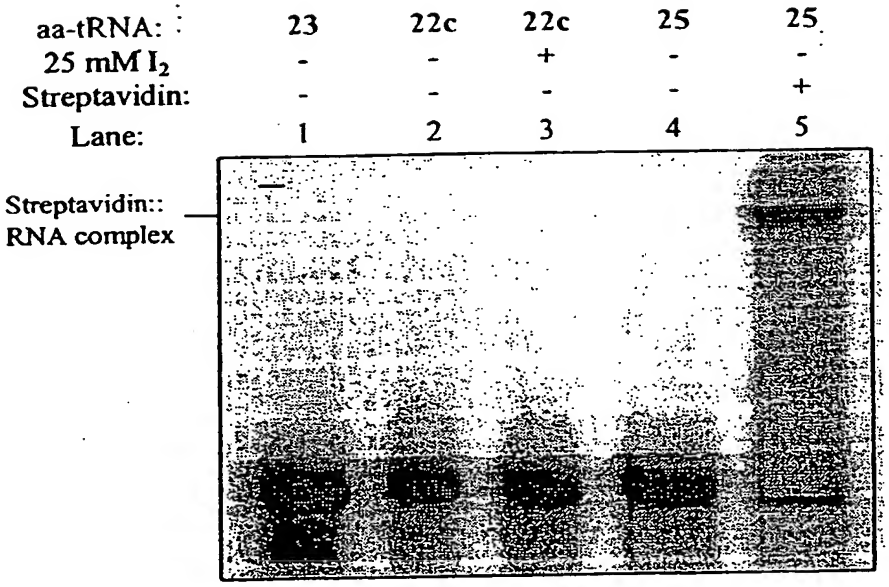
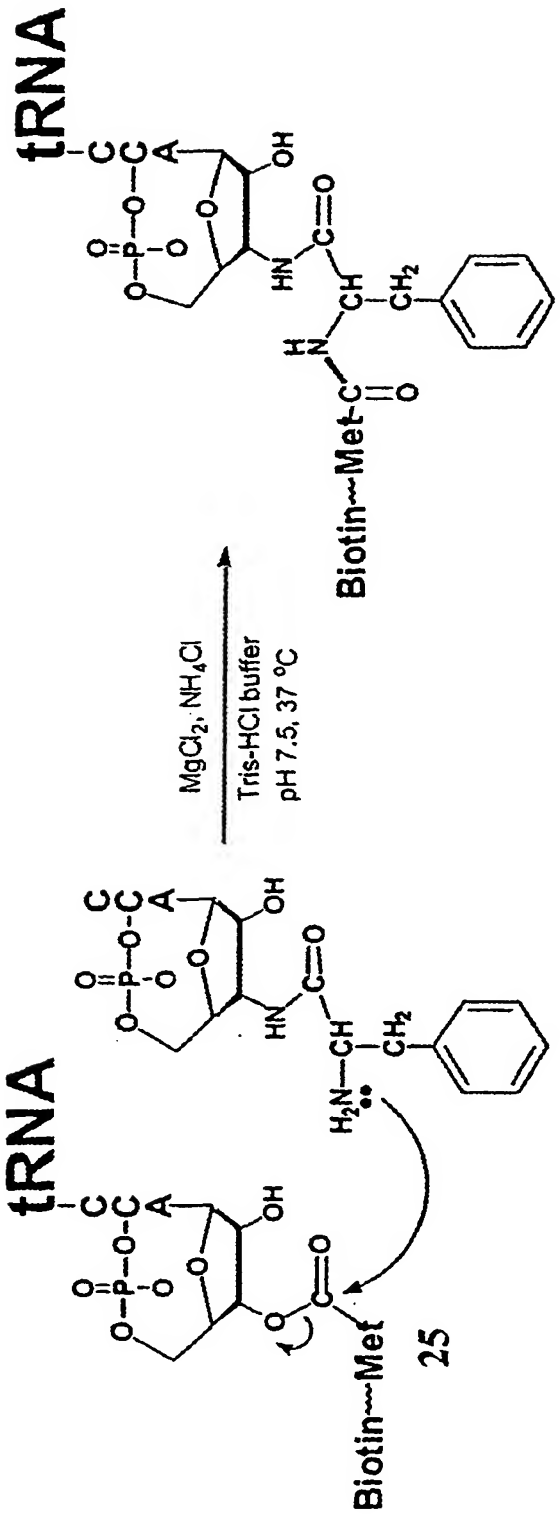
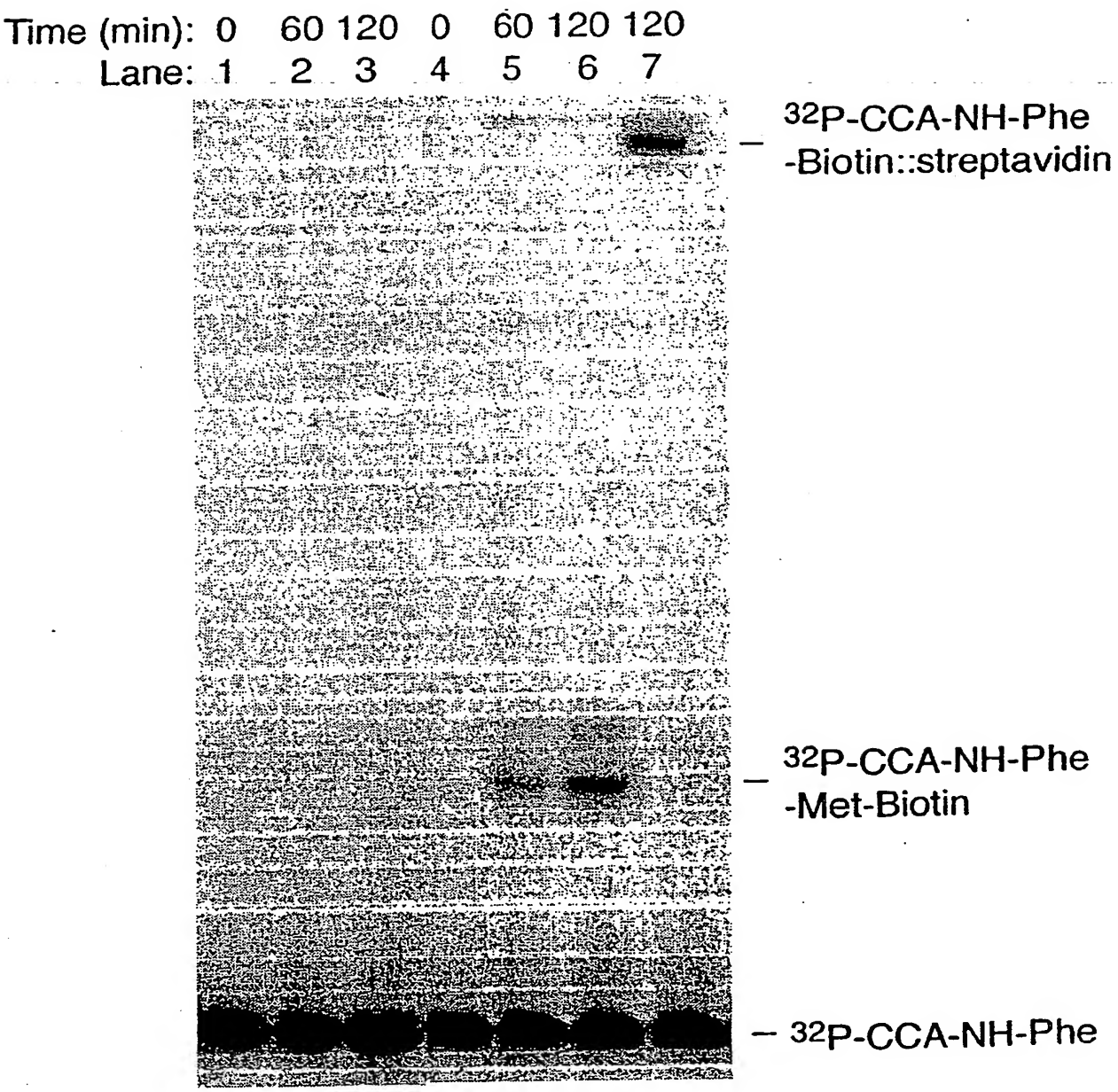

aa-tRNA: 23 22c 22c 25 25
25 mM I_2 - - + - -
Streptavidin: - - - - +
Lane: 1 2 3 4 5
Streptavidin:: RNA complex

Figure 6



BNSDOCID. <WO_____03059926A2_I_>

Figure 7



BNSDOCID: <WO_____03059926A2_I_>

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 July 2003 (24.07.2003)

PCT

(10) International Publication Number
WO 03/059926 A3

(51) International Patent Classification[7]: C07H 21/00

(21) International Application Number: PCT/US02/41064

(22) International Filing Date:
20 December 2002 (20.12.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/343,353 21 December 2001 (21.12.2001) US

(71) Applicant *(for all designated States except US)*: **UNIVERSITY OF MASSACHUSETTS** [US/US]; 365 Plantation Street, Worcester, MA 01605 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **ZHANG, Biliang** [CN/US]; 19 Faulkner Road, Shrewsbury, MA 01545 (US). **CUI, Zhiyong** [CN/US]; 65 Frank Street, Apartment #145, Worcester, MA 01604 (US).

(74) Agents: **GEARY, William, C., III** et al.; Nutter McClennen & Fish LLP, World Trade Center West, 155 Seaport Boulevard, Boston, MA 02210-2604 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

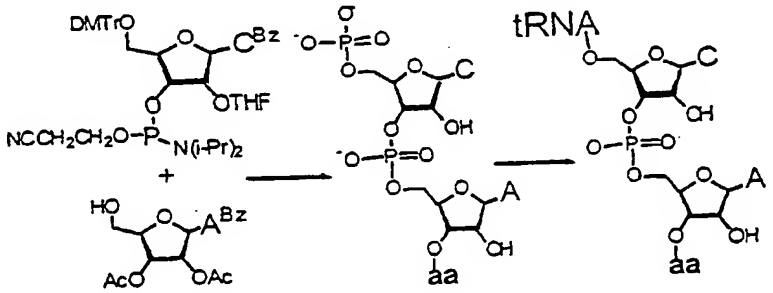
(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:
28 August 2003

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND COMPOSITIONS FOR AMINOACYL-tRNA SYNTHESIS



(57) Abstract: Methods and compositions are disclosed for the high yield chemical synthesis of 2'(3')-*O*-aminoacylated oligonucleotides, 2'(3')-*O*-aminoacylated pCpA derivatives and 2'(3')-aminoacyl-tRNA's. The present invention discloses the use of tetrahydrofuranyl group (THF) as a stable protecting group in the production of aminoacyl-oligonucleotides. THF can also be used in conjunction with a removable protecting group, such as dimethoxytrityl group (DMTr). The mild conditions employed for the removal of the THF group are compatible with the integrity of the aminoacyl linkage as well as tRNA, which makes it possible to utilize the methods of the present invention to form aminoacyl-tRNA's. The present invention discloses an efficient route for synthesizing 2'(3')-aminoacyl-pCpA, which can be used to load any natural amino acid, unnatural amino acid, labeled amino acid, reporter group or derivative thereof onto a tRNA molecule through an aminoacyl linkage.

WO 03/059926 A3

INTERNATIONAL SEARCH REPORT

Int· onal Application No
PCT/US 02/41064

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07H21/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, CHEM ABS Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	T. NIINOMI, M. SISIDO: "Synthesis of 2'(3')-O-aminoacyl-pdCpA carrying a photofunctional nonnatural amino acid" CHEMISTRY LETTERS, 1993, pages 1305-1308, XP002245631 scheme 1 ---	1-20
A	ROBERTSON S A ET AL: "A GENERAL AND EFFICIENT ROUTE FOR CHEMICAL AMINOACYLATION OF TRANSFER RNAS" JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, AMERICAN CHEMICAL SOCIETY, WASHINGTON, DC, US, vol. 113, no. 7, 1991, pages 2722-2729, XP001064836 ISSN: 0002-7863 the whole document ---	12
	-/--	

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

° Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier document but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
26 June 2003	11/07/2003

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	de Nooy, A

Form PCT/ISA/210 (second sheet) (July 1992)

BNSDOCID: <WO_____03059926A3_I_>

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 02/41064

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	NOREN C J ET AL: "A GENERAL METHOD FOR SITE-SPECIFIC INCORPORATION OF UNNATURAL AMINOACIDS INTO PROTEIN" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, US, vol. 244, 14 April 1989 (1989-04-14), pages 182-188, XP002925217 ISSN: 0036-8075 the whole document	12

BNSDOCID: <WO_____03059926A3_I_>

THIS PAGE BLANK (USPTO)